IJCSIS Vol. 9 No. 11, November 2011
ISSN 1947-5500

International Journal of Computer Science & Information Security

Editorial
Message from Managing Editor

***IJCSIS** editorial board consists of several internationally recognized experts and guest editors. Wide circulation is assured because libraries and individuals, worldwide, subscribe and reference to **IJCSIS**. The Journal has grown rapidly to its currently level of over thousands articles published and indexed; with distribution to librarians, universities, research centers, researchers in computing, and computer scientists. After a very careful reviewing process, the editorial committee accepts outstanding papers, among many highly qualified submissions (Acceptance rate below 30%). All submitted papers are peer reviewed and accepted papers are published in the IJCSIS proceeding **(ISSN 1947-5500).** The year 2011 has been very eventful and encouraging for all IJCSIS authors/researchers and IJCSIS technical committee, as we see more and more interest in IJCSIS research publications. IJCSIS is now empowered by over thousands of academics, researchers, authors/reviewers/students and research organizations. Reaching this milestone would not have been possible without the support, feedback, and continuous engagement of our authors and reviewers.*

The journal covers the frontier issues in the engineering and the computer science and their applications in business, industry and other subjects. (See monthly Call for Papers)

Since 2009, ***IJCSIS** is published using an open access publication model, meaning that all interested readers will be able to freely access the journal online without the need for a subscription. On behalf of the editorial committee, I would like to express my sincere thanks to all authors and reviewers for their great contribution.*

We wish everyone a successful scientific research year on 2012.

Available at http://sites.google.com/site/ijcsis/

IJCSIS Vol. 9, No. 11, November 2011 Edition

ISSN 1947-5500 © IJCSIS, USA.

Journal Indexed by (among others):

Google scholar

ScientificCommons

CiteSeerx beta

SCIRUS search engine for science

DOAJ DIRECTORY OF OPEN ACCESS JOURNALS

INDEX COPERNICUS INTERNATIONAL

ProQuest Start here.

IJCSIS EDITORIAL BOARD

TABLE OF CONTENTS

A Study of Elliptic Curves's Implementations Suitable for Embedded Systems

Moncef Amara #1 and Amar Siad #

# *LAGA Laboratory, University of Paris 8 (Vincennes Saint-Denis)*
Saint-Denis / FRANCE.
1 amara_moncef@yahoo.fr
1 moncef.amara02@etud.univ-paris8.fr

***Abstract*—The Elliptic Curve Cryptography (ECC) covers all relevant asymmetric cryptographic primitives like digital signatures and key agreement algorithms. ECC is considered as the best candidate for Public-Key Cryptosystems. Recently, Elliptic Curve Cryptography based on Binary Edwards Curves (BEC) has been proposed and it shows several interesting properties, e.g., completeness and security against certain exceptional-points attacks. In this paper, we present a study of the different methods to implement ECC in hardware, we study the implementation of the BEC to make it suitable for programmable devices, and we given as application a hardware design of elliptic curve operations over binary Fields $GF(2^m)$. The function used for this purpose is the scalar multiplication kP which is the core operation of ECCs. Where k is an integer and P is a point on an elliptic curve.**

***Index Terms*—Cryptography, Elliptic curves, Binary Edwards curve, Scalar multiplication, Binary arithmetic, Cryptosystems, Programmable devices, FPGA.**

I. Introduction

Elliptic Curve Cryptography (ECC) is a relatively new cryptosystem, suggested independently, from the second half oh 19th century, by Neals Koblitz [6] and Victor Miller [7]. At present, ECC has been commercially accepted, and has also been adopted by many standardizing bodies such as ANSI, IEEE, ISO and NIST [2]. Since then, it has been the focus of a lot of attention and gained great popularity due to the same level of security they provide with much smaller key sizes than conventional public key cryptosystems have.

The ECC covers all relevant asymmetric cryptographic primitives like digital signatures (ECDSA), key exchange and agreement protocols (ECDH). Point multiplication serves as the basic building block in all ECC primitives and is the computationally most expensive operation.

The best known and most commonly used public-key cryptosystems are RSA [8] and Elliptic Curve Cryptography (ECC) [7], [6]. The main benefit of ECC is that it offers equivalent security as RSA for much smaller parameter sizes. These advantages result in smaller data-paths, less memory usage and lower power consumption. ECC is widely considered as the best candidate for embedded systems.

Integrating a Public Key Cryptosystem into a embedded systems such as ASIC, FPGA and RFID-tag is a challenge due to the limitations in costs, area and power. On the other hand, security is required, in particular to prevent cloning or tracing. It was widely believed that devices with such constrained resources cannot carry out strong cryptographic operations such as Elliptic Curve Scalar Multiplication (ECSM). However, the feasibility of integrating PKCs into such devices have been recently proven by several implementations.

Standard formulas for adding two points, say P and Q, on a Weierstrass-form elliptic curves fail if P is at infinity, or if Q is at infinity, or if P+Q is at infinity. Binary Edwards curves provides a different equation to define an Elliptic Curve which no longer has points at infinity [1]. This feature is known as completeness.

The aim of this work is to present a study of state of the art of the different methods to implement ECC in hardware, intended to the conception of the hardware cryptographic applications. We present a complete study of binary Edwards curves to make it suitable for programmable devices, and we given a hardware design of elliptic curve operations over binary Fields $GF(2^m)$.

The paper is organized as follows. After a brief introduction, an overview of the use of elliptic curve in cryptography application is given in section 2. The point multiplication method is explained in Section 3, and binary Edwards curves are presented in Section 4. The EC Point multiplication processor given in Section 5. Finally, conclusion and open problems are summarized in Section 6.

II. Elliptic Curve Cryptography

Elliptic Curves, Fig.1, defined over a finite-field provide a group structure that is used to implement the cryptographic schemes. The elements of the group are the rational points on the elliptic curve, together with a special point O (called the "point at infinity").

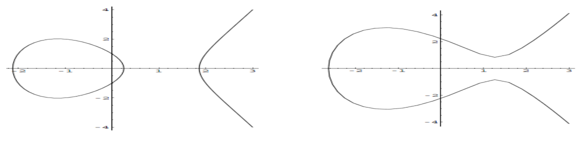


Fig. 1. Graphs of elliptic curves $y^2 = x^3 - 4x + 1$ (on the left) and $y^2 = x^3 - 5x + 5$ (on the right) over $\mathbb{R}$.

A major building block of all elliptic curve cryptosystems is the scalar point multiplication, an operation of the form $k.P$ where k is a positive integer and P is a point on the elliptic curve. Computing $k.P$ means adding the point P exactly $k-1$ times to itself, which results in another point Q on the elliptic curve. The inverse operation, i.e., to recover k when the points P and $Q = k.P$ are given, is known as the *Elliptic Curve Discrete Logarithm Problem* (ECDLP). To date, no subexponential-time algorithm is known to solve the ECDLP in a properly selected elliptic curve group. This makes Elliptic Curve Cryptography a promising branch of public key cryptography which offers similar security to other "traditional" DLP-based schemes in use today, with smaller key sizes and memory requirements, e.g., 160 bits instead of 1024 bits

A. Elliptic Curves over $\mathbb{F}_{2^m}$

In this section, a group operations on elliptic curves over $\mathbb{F}_{2^m}$ is described. A non-supersingular elliptic curve E over $\mathbb{F}_{2^m}$, $E(\mathbb{F}_{2^m})$ is the set of all solutions to the following equation [5]:

$$y^2 + xy = x^3 + a_2x^2 + a_6 \tag{1}$$

where $a_2, a_6 \in \mathbb{F}_{2^m}$, and $a_6 \neq 0$. Such an elliptic curve is a finite abelian group. The number of points in this group is denoted by $\#(E(\mathbb{F}_{2^m}))$.

1) ***Curve Addition****:* If $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ are points on the elliptic curve [i.e., satisfy (1)] and $P \neq -Q$, then $(x_3, y_3) = R = P + Q$ can be defined geometrically, Fig.2.

In the case that $P \neq Q$ (i.e., point addition), a line intersecting the curve at points P and Q and must also intersect the curve at a third point $-R = (x_3, -y_3)$.

2) ***Curve Doubling****:* If $P = Q$ (point doubling), the tangent line is used, Fig.3.

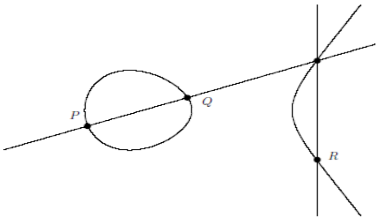


Fig. 2. Group law of elliptic curve (Point Addition).

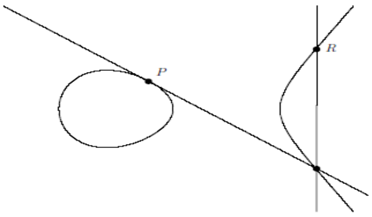


Fig. 3. Group law of elliptic curve (Point Doubling).

For E given in affine coordinates:

if $P \neq Q$:

$$\begin{aligned} x_3 &= \lambda^2 + \lambda + x_1 + x_2 + a \\ y_3 &= \lambda(x_1 + x_3) + x_3 + y_1 \\ \text{où } \lambda &= \frac{(y_2+y_1)}{(x_2+x_1)} \end{aligned} \tag{2}$$

if $P = Q$:

$$\begin{aligned} x_3 &= \lambda^2 + \lambda + a \\ y_3 &= x_1^2 + (\lambda+1)x_3 \\ \text{où } \lambda &= x_1 + \frac{y_1}{x_1} \end{aligned} \tag{3}$$

III. Elliptic Curve Point Multiplication

There are different ways to implement point multiplication: binary, signed digit representation (NAF), Montgomery method,. . ., etc. A scalar multiplication is performed in three different stages, Fig.4. At the top level, the method for computing the scalar multiplication must be selected, in the second level, the coordinates to represent elliptic points must be defined. From this representation, the Add operation is defined. Possible coordinates are : affine, projective, Jacobeans and L'opez-Dahab. The lower level, but the most important, involves the primitive field operations on which the curve is defined. Basic field operations are sum, multiplication, squaring and division.

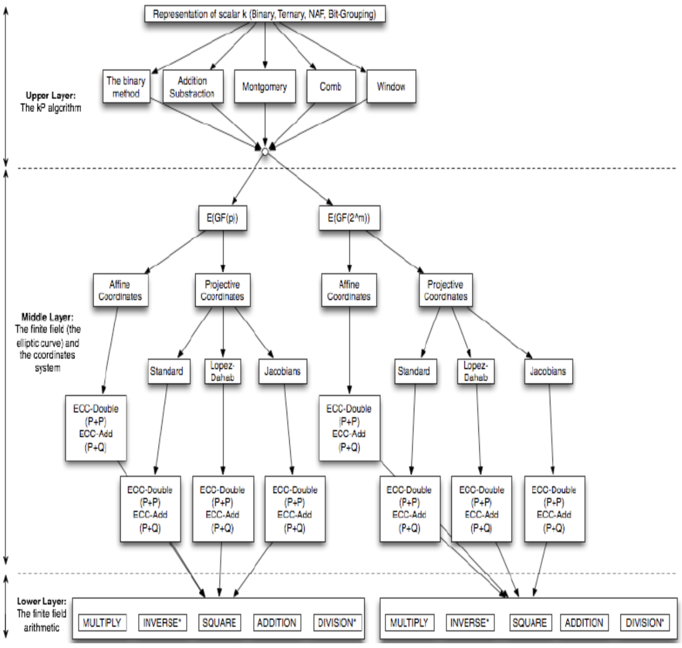

Representation of scalar k (Binary, Ternary, NAF, Bit-Grouping)
Upper Layer:
The kP algorithm
The binary method
Addition Substraction
Montgomery
Comb
Window
E(GF(p))
E(GF(2^m))
Middle Layer:
The finite field (the elliptic curve) and the coordinates system
Affine Coordinates
Projective Coordinates
Affine Coordinates
Projective Coordinates
Standard
Lopez-Dahab
Jacobians
Standard
Lopez-Dahab
Jacobians
ECC-Double (P+P) ECC-Add (P+Q)
ECC-Double (P+P) ECC-Add (P+Q)
ECC-Double (P+P) ECC-Add (P+Q)
ECC-Double (P+P) ECC-Add (P+Q)
ECC-Double (P+P) ECC-Add (P+Q)
ECC-Double (P+P) ECC-Add (P+Q)
ECC-Double (P+P) ECC-Add (P+Q)
ECC-Double (P+P) ECC-Add (P+Q)
Lower Layer:
The finite field arithmetic
MULTIPLY
INVERSE*
SQUARE
ADDITION
DIVISION*
MULTIPLY
INVERSE*
SQUARE
ADDITION
DIVISION*

Fig. 4. Different method to compute scalar multiplication $k.P$

A. Binary Method

The most simplest and straightforward implementation is the binary method, as shown in Algorithm.1. The binary method scans every bit of scalar k and, depending on its value, 0 or 1, it performs an ECC-DOUBLE operation or both a ECC-DOUBLE and an ECC-ADD operation. Algorithm.1, scans every bit of k from right to left.

For an elliptic curve defined on $\mathbb{F}_{2^m}$ using affine coordinates, the operations ECC-ADD and ECC-DOUBLE are performed according to equations (2) and (3) respectively. The operation ECC-ADD requires one inversion, two multiplications, one squaring and eight additions. The operation ECC-DOUBLE requires five additions, two squaring, two multiplications and one inversion, all of them, operations on $\mathbb{F}_{2^m}$.

Algorithm 1 *Binary method: right to left* [5]

```
Input: $P(x,y)$, $x,y \in GF(2^m)$, $k = (k_{m-1}, k_{m-2}, \ldots, k_0)$
Output: $R = k.P$
R ← 0
S ← P
for $i \leftarrow 0, m-1$ do
    if $k_i = 1$ then
        if $R = 0$ then
            $R \leftarrow S$
        else
            $R \leftarrow R + S$
        end if
    end if
    $S \leftarrow 2S$
end for
return $R$
```

B. Coordinates Systems

Table.I, summarizes the properties of the different coordinates systems; affine, projective, Jacobeans,..., etc. It should be noted that in all the cases the opposite of the point $(X : Y : Z)$ is written $(X : -Y : Z)$.

TABLE I
TABLE SUMMARIZING THE PROPERTIES OF THE VARIOUS PROJECTIVE COORDINATES SYSTEMS.

Coordinates	$(x,y) =$	Curve equation
$\mathcal{P}$	$(X/Z, Y/Z)$	$Y^2Z = X^3 + aXZ^2 + bZ^3$
$\mathcal{J}$	$(X/Z^2, Y/Z^3)$	$Y^2 = X^3 + aXZ^4 + bZ^6$
$\mathcal{J}^m$	$(X/Z^2, Y/Z^3)$	$Y^2 = X^3 + aXZ^4 + bZ^6$

The choice of the coordinate system is determined by the number of modular operations to carry out to calculate the doubling and the addition of points. Table.II, compares the cost of the doubling and the addition for each projective coordinate.

TABLE II
COST OF THE DOUBLING AND THE ADDITION FOR EACH PROJECTIVE COORDINATES SYSTEMS.

Coordinates	Cost of Double operation	Cost of Add operation
$\mathcal{A}$	$I + 4M$	$I + 3M$
$\mathcal{P}$	$12M$	**14M**
$\mathcal{J}$	$10M$	$16M$
$\mathcal{J}^m$	**8M**	$19M$

IV. EDWARDS CURVES

A new form for elliptic curves was added to the mathematical literature with Edwards curves. Edwards showed in [3] that all elliptic curves over number fields can be transformed to $x^2 + y^2 = c^2(1 + x^2y^2)$, with $(0, c)$ as the neutral element and with a simple and a symmetric addition law:

$$(x_1, y_1), (x_2, y_2) \rightarrow (\frac{x_1y_2 + y_1x_2}{c(1 + x_1x_2y_1y_2)} \frac{y_1y_2 + x_1x_2}{c(1 - x_1x_2y_1y_2)}) \tag{4}$$

A. Binary Edwards Curves

This section contains complete addition formulas for binary elliptic curves, i.e., addition formulas that work for all input pairs, with no exceptional cases. First, the need for Edwards curves is explained, and then the theorems and formulas will be shown in order.

The points on a Weierstrass-form elliptic curve:

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 \tag{5}$$

include not only the affine point (x_1, y_1), but also an extra point at infinity serving as neutral element. The standard formulas for elliptic curve to compute a sum $P_1 + P_2$ fail if P_1, P_2, or $P_1 + P_2$ is at infinity, or if P_1 is equal to P_2. Each of these possibilities should be tested separately before generating any elliptic curve cryptosystem.

Definition 1: (Binary Edwards Curve) Let k be a field with $char(k) = 2$. Let d_1, d_2 be elements of k with $d_1 \neq 0$ and $d_2 \neq d_1^2 + d_1$, then the binary Edwards curve with coefficients d_1 and d_2 is the affine curve:

$$E_{B,d_1,d_2} = d_1(x+y) + d_2(x^2+y^2) = xy + xy(x+y) + x^2y^2 \tag{6}$$

This curve is symmetric in x and y and thus it has the property that if (x_1, y_1) is a point on the curve then so is (y_1, x_1). The point $(0, 0)$ will be the neutral element of the addition law, while $(1, 1)$ will have order 2.

B. Binary Edwards Curves Addition Law

Binary Edwards curves, E_{B,d_1,d_2}, addition law is given as in follows, and it is proven that the addition law corresponds to the elliptic curve in Weierstrass form similarly. It can be used for doubling with two identical inputs. The sum of two points $(x_1, y_1), (x_2, y_2)$ on E_{B,d_1,d_2} is the point (x_3, y_3) defined as follows:

$$x_3 = \frac{d_1(x_1+x_2)+d_2(x_1+y_1)(x_2+y_2)+(x_1+x_1^2)(x_2(y_1+y_2+1)+y_1y_2)}{d_1+(x_1+x_1^2)(x_2+y_2)} \tag{7}$$

$$y_3 = \frac{d_1(y_1+y_2)+d_2(x_1+y_1)(x_2+y_2)+(y_1+y_1^2)(y_2(x_1+x_2+1)+x_1x_2)}{d_1+(y_1+y_1^2)(x_2+y_2)} \tag{8}$$

If the denominators:

$$d_1 + (x_1 + x_1^2)(x_2 + y_2)$$

and

$$d_1 + (y_1 + y_1^2)(x_2 + y_2)$$

are non-zero then the sum (x_3, y_3) is a point on E_{B,d_1,d_2}: i.e.,

$$d_1(x_3 + y_3) + d_2(x_3^2 + y_3^2) = x_3.y_3 + x_3.y_3(x_3 + y_3) + x_3^2.y_3^2$$

Here, if the points are inserted like $(0, 0)$ into the addition law, it is shown that $(0, 0)$ is the neutral element. Similarly,

$(x_1, y_1) + (1,1) = (x_1 + 1, y_1 + 1)$; in particular $(1,1) + (1,1) = (0,0)$. Furthermore $(x_1, y_1) + (y_1, x_1) = (0,0)$, so $-(x_1, y_1) = (y_1, x_1)$.

C. Explicit Addition Formulas

In this section, we present explicit formulas for affine addition, projective addition on the binary Edwards curves.

*1) **Affine Addition**:* The following formulas, given (x_1, y_1) and (x_2, y_2) on the binary Edwards curve E_{B,d_1,d_2}, compute the sum $(x_3, y_3) = (x_1, y_1) + (x_2, y_2)$ if it is defined:

Algorithm 2 *Affine Addition*

```
$w_1 = x_1 + y_1,$
$w_2 = x_2 + y_2,$
$A = x_1^2 + x_1,$
$B = y_1^2 + y_1,$
$C = d_2 w_1 w_2,$
$D = x_2 y_2,$
$x_3 = y_1 + (C + d_1(w_1 + x_2) + A(D + x_2))/(d_1 + A w_2),$
$y_3 = x_1 + (C + d_1(w_1 + y_2) + B(D + y_2))/(d_1 + B w_2).$
```

These formulas use $2I + 8M + 2S + 3D$, where I is the cost of inversion, M is the cost of multiplication, S is the cost of squaring, D is the cost of a multiplication by a curve parameter. The $3D$ here are two multiplications by d_1 and one multiplication by d_2 [1].

*2) **Projective Addition**:* The following formulas, given $(X_1 : Y_1 : Z_1)$ and $(X_2 : Y_2 : Z_2)$ on the binary Edwards curve E_{B,d_1,d_2}, compute the sum $(X_3 : Y_3 : Z_3) = (X_1 : Y_1 : Z_1) + (X_2 : Y_2 : Z_2)$.

Algorithm 3 *Projective Addition*

```
$W_1 = X_1 + Y_1,$
$W_2 = X_2 + Y_2,$
$A = X_1.(X_1 + Z_1),$
$B = Y_1.(Y_1 + Z_1),$
$C = Z_1.Z_2,$
$D = W_2.Z_2,$
$E = d_1.C.C,$
$F = (d_1 Z_2 + d_2 W_2).W_1.C,$
$G = d_1.C.Z_1,$
$U = E + A.D,$
$V = E + B.D,$
$S = U.V,$
$X_3 = S.Y_1 + (F + X_2(G + A(Y_2 + Z_2))).V.Z_1,$
$Y_3 = S.X_1 + (F + Y_2(G + B(X_2 + Z_2))).U.Z_1,$
$Z_3 = S.Z_1.$
```

These formulas use $21M + 1S + 4D$. The $4D$ are three multiplications by d_1 and one multiplication by d_2.

D. Binary Edwards Curves Doubling Law

The doubling formulas on the Edwards curve E_{B,d_1,d_2} is presented in this section. Affine coordinates and inversion-free projective coordinates are given respectively.

*1) **Affine Doubling**:* Let (x_1, y_1) be a point on E_{B,d_1,d_2}, and assume that the sum $(x_1, y_1) + (x_1, y_1)$ is defined. Computing $(x_3, y_3) = (x_1, y_1) + (x_1, y_1)$ we obtain:

$$\begin{aligned} x_3 &= \tfrac{d_1(x_1+y_1)^2+(x_1+x_1^2)(x_1+y_1^2)}{d_1+(x_1+y_1)(x_1+x_1^2)} \\ &= \tfrac{d_1(x_1+y_1)+x_1y_1+x_1^2(1+x_1+y_1)}{d_1+x_1y_1+x_1^2(1+x_1+y_1)} \\ &= 1 + \tfrac{d_1(1+x_1+y_1)}{d_1+x_1y_1+y_1^2(1+x_1+y_1)} \end{aligned} \tag{9}$$

Also we obtain:

$$y_3 = 1 + \frac{d_1(1 + x_1 + y_1)}{d_1 + x_1 y_1 + y_1^2(1 + x_1 + y_1)} \tag{10}$$

Note that, the affine formulas is computed with one inversion, as the product of the denominators of x_3 and y_3 is:

$$(d_1 + x_1 y_1 + x_1^2(1 + x_1 + y_1))(d_1 + x_1 y_1 + y_1^2(1 + x_1 + y_1))$$

$$\begin{aligned} &= d_1^2 + (x_1^2 + y_1^2)(d_1(1 + x_1 + y_1) + x_1 y_1(1 + x_1 + y_1) + x_1^2 y_1^2) \\ &= d_1^2 + (x_1^2 + y_1^2)(d_1 + d_2(x_1^2 + y_1^2)) \\ &= d_1(d_1 + x_1^2 + y_1^2 + (d_2/d_1)(x_1^4 + y_1^4)) \end{aligned} \tag{11}$$

where the curve equation is used again. This leads to the doubling formulas:

$$x_3 = 1 + \frac{d_1 + d_2(x_1^2 + y_1^2) + y_1^2 + y_1^4}{d_1 + x_1^2 + y_1^2 + (d_2/d_1)(x_1^4 + y_1^4)} \tag{12}$$

$$y_3 = 1 + \frac{d_1 + d_2(x_1^2 + y_1^2) + x_1^2 + x_1^4}{d_1 + x_1^2 + y_1^2 + (d_2/d_1)(x_1^4 + y_1^4)} \tag{13}$$

which needs $1I + 2M + 4S + 2D$.
If $d_1 = d_2$ some multiplications can be grouped as follows:

Algorithm 4 *Affine Doubling*

```
$A = x_1^2,$
$B = A^2,$
$C = y_1^2,$
$D = C^2,$
$E = A + C,$
$F = 1/(d_1 + E + B + D),$
$x_3 = (d_1 E + A + B).F,$
$y_3 = x_3 + 1 + d_1 F.$
```

These formulas use only $1I + 1M + 4S + 2D$.

*2) **Projective Doubling**:* In this sub-section, explicit formulas of projective doubling is given to compute $2(X_1 : Y_1 : Z_1) = (X_3 : Y_3 : Z_3)$:

Algorithm 5 *Projective Doubling*

```
$A = X_1^2$,
$B = A^2$,
$C = Y_1^2$,
$D = C^2$,
$E = Z_1^2$,
$F = d_1.E^2$,
$G = (d_2/d_1).(B + D)$,
$H = A.E$,
$I = C.E$,
$J = H + I$,
$K = G + d_2.J$,
$X_3 = K + H + D$,
$Y_3 = K + I + B$,
$Z_3 = F + J + G$.
```

These formulas use $2M + 6S + 3D$. The $3D$ are multiplications by $d_1, d_2/d_1$ and d_2.

V. An Application of Elliptic Curve Implementation over $GF(2^m)$

A. Field Programmable Gate Array (FPGA)

Field programmable gate array (FPGA) devices provide an excellent technology for the implementation of general purpose cryptographic devices. Compared with application specific integrated circuits (ASIC), FPGA as offer low nonrecurring engineering costs, shorter design time, greater flexibility and the ability to change the algorithm or design.

Fig.5, shows a structure of ECC processor. It consists of a main control block, an ECC add and double block and an ECC block for arithmetic operations. The ECC processor we have implemented is defined over the field $GF(2^{163})$, which is a SEC-2 recommendation [9], with this field being defined by the field polynomial $F(x) = x^{163}+x^7+x^6+x^3+1$.

The EC point multiplication processor, defined in affine coordinates, is achieved by using a dedicated Galois Field arithmetic, implemented on FPGA using VHDL language.

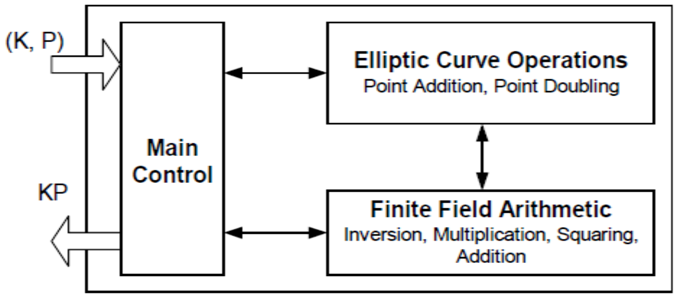

(K, P)
KP
Main Control
Elliptic Curve Operations
Point Addition, Point Doubling
Finite Field Arithmetic
Inversion, Multiplication, Squaring, Addition

Fig. 5. Elliptic curve point multiplication processor.

Fig.6, shows the hardware implementation of point addition operation, corresponding to equation (2).

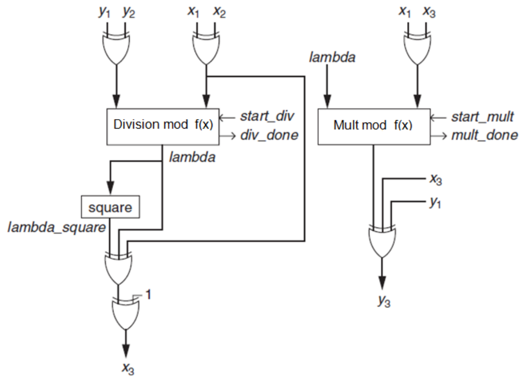


Fig. 6. Hardware implementation of point addition operation.

Multiplication in $GF(2^m)$ with polynomial basis representation is presented in this section. Inputs $A = (a_0, a_1, \ldots, a_{m-1})$ and $B = (b_0, b_1, \ldots, b_{m-1}) \in GF(2^m)$, and the product $C = AB = (c_0, c_1, \ldots, c_{m-1})$ are treated as polynomials $A(x), B(x)$, and $C(x)$ with respective coefficients. The dependence between these polynomials is given by $C(x) = A(x).B(x) \bmod F(x)$, Where $F(x)$ is a constant irreducible polynomial of degree m. The hardware implementation for multiplication in $GF(2^m)$ is presented in Fig.7.

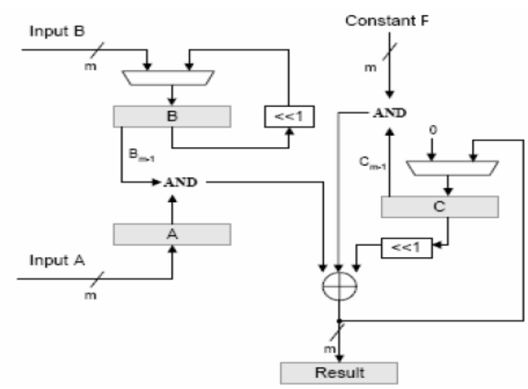


Fig. 7. Serial Multiplier in $GF(2^m)$.

The Hardware implementation of inversion in $GF(2^m)$ is presented in Fig.8.

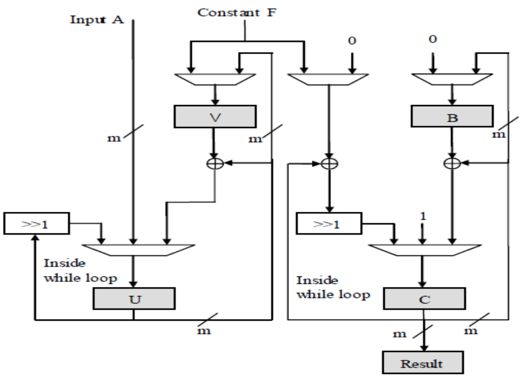


Fig. 8. Inverter in $GF(2^m)$.

B. Simulation and Results: The use of NIST-Recommended Elliptic Curves

The NIST elliptic curves over $\mathbb{F}_{2^{163}}$ and $\mathbb{F}_{2^{233}}$ are listed in Table.II. The following notation is used. The elements of $\mathbb{F}_{2^m}$ are represented using a polynomial basis representation with reduction polynomial $f(x)$. The reduction polynomials for the fields $\mathbb{F}_{2^{163}}$ and $\mathbb{F}_{2^{233}}$ are $f(x) = x^{163}+x^7+x^6+x^3+1$ and $f(x) = x^{233}+x^{74}+1$ respectively. An elliptic curve E over $\mathbb{F}_{2^m}$ is specified by the coefficients $a, b \in \mathbb{F}_{2^m}$ of its defining equation $y^2 + xy = x^3 + ax^2 + b$. The number of points on E defined over $\mathbb{F}_{2^m}$ is nh, where n is prime, and h is called the co-factor. A random curve over $\mathbb{F}_{2^m}$ is denoted by B-m.

TABLE III
NIST-RECOMMENDED ELLIPTIC CURVES OVER $\mathbb{F}_{2^{163}}, \mathbb{F}_{2^{233}}$ [4].

B-163:	$m = 163, f(z) = z^{163} + z^7 + z^6 + z^3 + 1,$ $a = 1, h = 2$
b	= 0x 00000002 0A601907 B8C953CA 1481EB10 512F7874 4A3205FD
n	= 0x 00000004 00000000 00000000 000292FE 77E70C12 A4234C33
x	= 0x 00000003 F0EBA162 86A2D57E A0991168 D4994637 E8343E36
y	= 0x 00000000 D51FBC6C 71A0094F A2CDD545 B11C5C0C 797324F1
B-233:	$m = 233, f(z) = z^{233} + z^{74} + 1,$ $a = 1, h = 2$
b	= 0x 00000066 647EDE6C 332C7F8C 0923BB58 213B333B 20E9CE42 81FE115F 7D8F90AD
n	= 0x 00000100 00000000 00000000 00000000 0013E974 E72F8A69 22031D26 03CFE0D7
x	= 0x 000000FA C9DFCBAC 8313BB21 39F1BB75 5FEF65BC 391F8B36 F8F8EB73 71FD558B
y	= 0x 00000100 6A08A419 03350678 E58528BE BF8A0BEF F867A7CA 36716F7E 01F81052

C. Implementation

For implementation, the architecture has been tested on ISE 9.2i Software using XILINX FPGA xc5vlx50-3-ff1153 device and simulate with ISE Simulator.

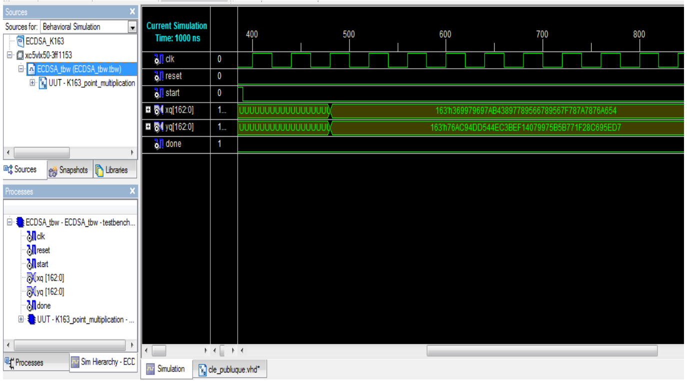

Sources
Sources for: Behavioral Simulation
ECDSA_K163
xc5vlx50-3ff1153
ECDSA_tbw (ECDSA_tbw.tbw)
UUT - K163_point_multiplication
Sources
Snapshots
Libraries
Processes
ECDSA_tbw - ECDSA_tbw - testbench...
clk
reset
start
xq [162:0]
yq [162:0]
done
UUT - K163_point_multiplication - ...
Processes
Sim Hierarchy - ECD
Current Simulation Time: 1000 ns
400
500
600
700
800
clk 0
reset 0
start 0
xq[162:0] 1...
yq[162:0] 1...
done 1
UUUUUUUUUUUUUUUUUU
163'h369979697AB438977895667895 67F787A7876A654
UUUUUUUUUUUUUUUUUU
163'h76AC94DD544EC3BEF14079975B5B771F28C695ED7
Simulation
cle_publuque.vhd*

Fig. 9. Simulation with ISE of scalar multiplication $k.P$ for $E(\mathbb{F}_{2^{163}})$

TABLE IV
THE x AND y INPUT COORDINATES OF THE POINT P AND AN ARBITRARY VALUE OF k.

k	= 0x 00000001 33E3CAE7 2CD0F448 B2954810 FB75B5E3 D8F43D07
P_x	= 0x 00000003 69979697 AB438977 89566789 567F787A 7876A654
P_y	= 0x 00000004 035EDB42 EFAFB298 9D51FEFC E3C80988 F41FF883

Table.3 shows the input parameters of the ECC scalar multiplication for a "163 bits" arbitrary value of k, and in Table.V, we give the implementation results corresponding.

TABLE V
SYNTHESIS RESULTS FOR $E(\mathbb{F}_{2^{163}})$.

point multiplication $G(F_{2^{163}})$		
Slice Logic Utilization:		
Number of Slice Registers:	2163	7%
Number of Slice LUTs:	2735	9%
Number used as Logic:	2735	9%
IO Utilization:		
Number of bonded IOBs:	330	58%
Maximum Frequency:	169.477MHz	

In Table.VI, we give the implementation results for $\mathbb{F}_{2^{233}}$.

TABLE VI
SYNTHESIS RESULTS FOR $E(\mathbb{F}_{2^{233}})$.

point multiplication $G(F_{2^{233}})$		
Slice Logic Utilization:		
Number of Slice Registers:	3073	10%
Number of Slice LUTs:	3637	12%
Number used as Logic:	3637	12%
IO Utilization:		
Number of bonded IOBs:	470	83%
Maximum Frequency:	136.323MHz	

VI. CONCLUSION AND OPEN PROBLEMS

In this work, the elliptic curve point multiplication is considered. we have presented the different methods which can be used to implement ECC in hardware, we have given an interesting study of the implementation of the Binary Edwards curves, and we have presented a version of an ECC crypto-hardware based on a Add and Double method, implemented on a Xilinx Virtex 5 device.

This study can be extended by developing a digital signature algorithm, which is very important in cryptography and internet security areas.

REFERENCES

[1] D.J. Bernstein, T. Lange and R.R. Farashahi. *Binary Edwards Curves*. Cryptology ePrint Archive, Report 2008/171, 2008, http://eprint.iacr.org/.

[2] Digital Signature Standard (DSS). Federal Information Processing Standards Publication 186-2, National Institute of Standards and Technology. 2000.

[3] H.M. Edwards. *A Normal Form for Elliptic Curves*. Bulletin of the American Mathematical Society, vol. 44, no. 3, pp. 393–422, July 2007.

[4] D. Hankerson, J. L'opez Hernandez and A. Menezes. *Software Implementation of Elliptic Curve Cryptography over Binary Fields*. In Proceedings of the Second International Workshop on Cryptographic Hardware and Embedded Systems (CHES), volume 1965 of Lecture Notes in Computer Science. 2001.

[5] D. Hankerson, A. Menezes and S. Vanstone. *Guide to Elliptic Curve Cryptography*. Springer, 2004.

[6] N. Koblitz. *Elliptic Curve Crytosystems*. Mathematics of Computation, Vol. 48, pages 203-209, 1987.

[7] V.S. Miller. *Use of Elliptic Curves in Cryptography*. Advances in Cryptology-CRYTO '85, Lecture Notes in Computer Science, vol. 128, Springer-Verlag, pages 417-426, 1985, Hugh C. Williams (Ed.).

[8] R.L. Rivest, A. Shamir and L.M. Adleman. *A Method for Obtaining Digital Signatures and Public-Key Cryptosystems*. Commun. ACM, vol. 21, no. 2, pp. 120–126, 1978.

[9] SEC 2: *Recommended Elliptic Curve Domain Parameters. Standard for Efficient Cryptography*. The SECG Group. 2000.

Transformation Invariance and Luster Variability in the Real-Life Acquisition of Biometric Patterns

R. Bremananth,
Information Systems and Technology Department,
Sur University College, Affiliated to Bond University, Australia
P.O. 440, Postal code 411,
Sur, Oman.
bremresearch@gmail.com / bremananth@suc.edu.om

Abstract— **In the real-life scenario, obtaining transformation invariant feature extraction is a challenging task in Computer Vision. Biometric recognitions are suffered due to diverse luster variations and transform patterns especially for face and biometric features. These patterns are main contingence on the distance of acquisition from the sensor to subjects' location and the external luster of the environments that make diverse revolutionizes in the biometric features. Another invariant aspect is the translation and rotation. Explicitly face and biometric features should be a positional independent whenever an Active-Region-of-Pattern (AROP) can occur anyplace in the acquired image. In this research paper, we propose Jacobin based transformation invariance scheme. The method is effectively incorporated in order to attain essential features which are required for the transformation invariant recognition. The results show that the proposed method can robust in the real-life Computer vision applications.**

Keywords- Biometric; Luster variations; Jacobian transformation; Transformation invariant patterns;

I. INTRODUCTION

Transformation invariant pattern recognition plays an essential role in the field of computer vision, pattern recognition, document analysis, image understanding and medical imaging. Since the system works well for the invariant real-life transformation distortions, it turns into an efficient recognition or identification system. In addition, features extracted from the identical sources should be classified as the same kind of classes in diverse luster and other deformation. An invariant pattern recognition system is capable of adjusting to any exterior artifacts and produces minimum false positives for the patterns that are obtained from the intra-classes. The aim of this paper is to suggest transformation invariant pattern recognition that improves the performance of recognition system. Images can be acquired either by a still camera or extracting frames from a motion sequence of video camera using a standard frame grabber or capture card. However, the latter method is more suitable for real-life processing because it produces sequence of images from which system can choose the best frame for the preprocessing. Image status checking has been carried out to select an image, which is capable for further processing such as binarization, localization and other recognition operations. Threshold analysis is an essential process to choose a set of minimum and maximum values for the real-life images, which are provided an efficient preprocessing in different kinds of luster. In the current literature, Image quality assessment was discussed by Li ma et al [1] to select a sharp Biometric image from the input sequence using Fourier transformation. However, there was no distance of capturing in between camera and subject position reported in the literature [2] [3].

Currently, biometric camera is capable to capture the eye images up to 36 inches with clear pigments of biometrics, though this paper analyses biometric images, which are acquired from 18 inches to 48 inches. Moreover, eye images are captured in divergent orientations with different luster and by varying distance between biometric camera and subjects that are challenges to the proposed methodology. Furthermore, Anti-spoofing module aims to allow living human beings by checking the pupil diameter variations in diverse luster at the same distance of capturing. It prevents artificial eye images to be enrolled or verified by the system. This method is known as challenge-response test.

Invariant feature extraction is a difficult problem in computer vision to recognize a person in non-invasive manner, for instance, from a long distance. It provides high security in any public domain application such as E-election, bank transactions, network login and other automatic person identification systems. The algorithm can be categorized into four types such as Quadrature-phasor encoded, Texture analysis, Zero-crossing, Local variation methods and rotation invariant feature extraction for biometrics were suggested by Daugman [3] , Li ma et al [1], Li ma et al [2], and Bremananth et al [4][5][6], respectively. However, these methods have limitations such as masking bits for occlusion avoidance, shifting of feature bits and several templates required to make a system as rotation invariant. Locating active-region-of-pattern (AROP) is complicated processes in the diverse environment and luster variations that include luster correction, invariance localization, segmentation, transformation invariant feature extraction and recognition.

The remainder of this paper is organized as follows: Section II emphasizes on transformation pattern extraction, geometric and luster transformation functions. Issues of transformation invariant pattern recognition are described in Section III. Section IV depicts the results obtained based on the proposed methodologies. Concluding remarks and future research direction are given Section V.

II. TRANSFORMATION PATTERNS EXTRACTION

The basic geometric transformations are usually employed in Computer Graphics and Visualization, and are often executed in Image analysis, Pattern recognition and Image understanding as well (Milan Sonka et al [7]). They allow exclusion of image deformations that occur when images are captured in a real-life condition. If one strives to match two different images of the same subject, an image transformation should be required to compensate their changes in orientation, size and shapes. For example, if one is trying to capture and match a remotely sensed eye images from the same area even after a minute, the recent image will not match exactly with the previous image due to factors such as position, scale, rotation and changes in the patina. To examine these alterations, it is necessary to execute an image transformation and then recognize the images. Skew occurs while capturing images with an obvious orientation at the diverse angles. These variations may be very tiny, but can be critical if the orientation is demoralized in subsequent processing. This is normally a problem in computer vision applications such as character, Biometric and license plate recognition.

The basic transformation is a vector function T that maps the pixel (x,y) to a new position (x',y') described as

$$x' = T_x(x, y) \qquad y' = T_y(x, y), \tag{1}$$

where T_x and T_y are transformation equations.

It transforms pixels into point-to-point basis. The commonly used transformations in recognition systems are pixel coordinate and brightness transformations. Pixel coordinate transformation is used to map the coordinate points of input pixel to a point in the output image. Figure 1 illustrates pixel coordinate transformation.

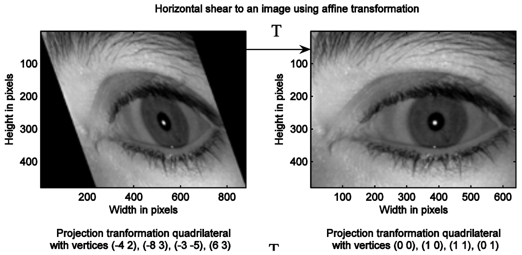

Horizontal shear to an image using affine transformation
T
Height in pixels
100
200
300
400
200
400
600
800
Width in pixels
Projection tranformation quadrilateral with vertices (-4 2), (-8 3), (-3 -5), (6 3)
T
Height in pixels
100
200
300
400
100
200
300
400
500
600
Width in pixels
Projection tranformation quadrilateral with vertices (0 0), (1 0), (1 1), (0 1)

Figure 1. Pixel coordinate transformation for biometric image Transformation on an image plane.

Equation (1) is usually approximated by the polynomial (Milan Sonka et al 1999) as shown below

$$x^{'} = \sum_{r=0}^{m}\sum_{k=0}^{m-r} a_{rk} x^r y^k, \qquad y^{'} = \sum_{r=0}^{m}\sum_{k=0}^{m-r} b_{rk} x^r y^k, \tag{2}$$

where a_{rk}, b_{rk} are linear coefficients, (x, y) is the known point and (x', y') is the transformed point in the output image. It is possible to determine a_{rk}, b_{rk} by solving the linear equations, if both coordinate points are known. When the geometric transform does not change rapidly depending on the position in the image lower order approximation polynomials (m = 2 or 3) are used with 6 -10 pairs of corresponding points. These points should be distributed in the image in such a way that it can articulate the geometric transformation. Typically corresponding points are spread uniformly. When the geometric transform is sensitive to the distribution of corresponding points in the input, higher degree of approximating polynomials are used. Equation (3) is approximately with four pairs of corresponding points by the bilinear transform described as

$$\begin{aligned} x' &= a_0 + a_1 x + a_2 y + a_3 xy \\ y' &= b_0 + b_1 x + b_2 y + b_3 xy. \end{aligned} \tag{3}$$

The affine transformation requires three pairs of corresponding points to find the coefficients as in (4). The affine transform includes geometric transformation such as rotation, translation, scaling and skewing.

$$\begin{aligned} x' &= a_0 + a_1 x + a_2 y \\ y' &= b_0 + b_1 x + b_2 y. \end{aligned} \tag{4}$$

A transformation applied to the entire image may alter the coordinate system. Jacobian J provides information about how the co-ordinates are modified due to the transformations. This is represented as

$$J = \left| \frac{\partial(x', y')}{\partial(x, y)} \right| = \begin{vmatrix} \dfrac{\partial x'}{\partial x} & \dfrac{\partial x'}{\partial y} \\ \dfrac{\partial y'}{\partial x} & \dfrac{\partial y'}{\partial y} \end{vmatrix}. \tag{5}$$

If transformation is singular J = 0. If the area of an image is invariant under the transformation then J = 1. The Jacobian for the bilinear and affine transform is described in (6) and (7), respectively.

$$J = a_1 b_2 - a_2 b_1 + (a_1 b_3 - a_3 b_1)x + (a_3 b_2 - a_2 b_3)y, \tag{6}$$

$$J = a_1 b_2 - a_2 b_1. \tag{7}$$

A. Geometric Transformations.

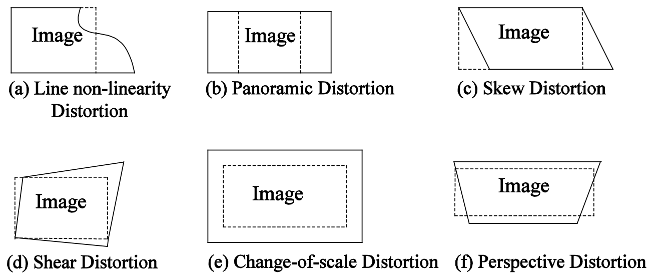
Biometric feature extraction depends on geometric data transformation. We can see that face and Biometric images have mainly rotation transformations. In the real-life scenario, patterns are acquired by the sensors, due to rotation, translation and scaling, they are notably diverge. So that any robust algorithms could be suffered to extract unique templates in order to obtain their prominent by nature. For example, Table I describes some of the various geometric transformations which could be occurred duration acquisition of biometric patterns. From these seven transformation types, we believe that biometric patterns could be adapted to the amply circumstances which are habitually transpired on their pattern catastrophe.

TABLE I. GEOMETRIC TRANSFORMATION FUNCTIONS.

No.	**Transformation Types**	**Transformation Function**	
		(x', y')	**J**
1	Rotation through an angle ϕ about the origin in clockwise direction	$x' = x\cos\phi + y\sin\phi$ $y' = x\sin\phi + y\cos\phi$	$J = 1$
2	Rotation through the angle ϕ about the origin in anticlockwise direction	$x' = x\cos\phi - y\sin\phi$ $y' = x\sin\phi + y\cos\phi$	$J = 1$
3	Rotation through the angle ϕ about rotation point (x_r,y_r) in anticlockwise direction	$x' = x_r + (x - x_r)\cos\varphi - (y - y_r)\sin\varphi$ $y' = y_r + (x - x_r)\sin\phi + (y - y_r)\cos\phi$	$J = 1$
4	Scaling a in x-axis and b in y-axis	$x' = ax \quad y' = bx$	$J = ab$
5	Fixed point scale	$x' = x_f + (x - x_f)a$ $y' = y_f + (y - y_f)b$	$J = ab$
6	Skew by the angle ϕ	$x' = x + y\tan\phi \quad y' = y$	$J = 1$
7	Translation	$x' = x + t_x \quad y' = y + t_y$	$J = 1$

Any complex upheaval can be approximated by partitioning an image into smaller rectangular sub-images. Image upheaval is estimated on the corresponding pair of pixels by using affine or bilinear method and then repairing each sub-image separately. An optical camera is a passive sensor, which offers more affordable non-linearities in raster scanning and a non-constant sampling period in capturing any moving object. There are some cataclysms that must be tackled in remote sensing. The main source of rotation, skew, scale, translation and non-linearity upheaval are due to the wrong position or orientation of the camera or sensor with respect to the object or diverse way of acquiring an object. Figure 2 shows some of the distortions that occur while capturing an object by any type of passive sensor.

Line non-linearity distortion is caused by variable distance of the object from the camera mirror as shown in Fig. 2a. Camera mirror rotating at constant speed causes panoramic parody. This is shown is Fig. 2b. The rotation or shake of an object during image capturing produces skew distortion as shown in Fig. 2c. The shear distortion is represented in Fig. 2d. The variation of distance between the object and camera provokes change-of-scale distortion as shown in Fig. 2e. Figure 2f shows the perspective distortions.



(a) Line non-linearity Distortion (b) Panoramic Distortion (c) Skew Distortion (d) Shear Distortion (e) Change-of-scale Distortion (f) Perspective Distortion

Figure 2. Types of upheaval occurred in the real-life acquisition.

B. Luster Transformation function

Luster transformation functions are principally for reimbursing sheens and gleams of picture elements which are be revealed on the acquisition of biometric patterns. Here, we programmed for some of the luster which are possibly devastatingly exaggerated on the patterns and features of the face and Biometric images. Table II listed some of the luster transformation functions.

TABLE II. LUSTER TRANSFORMATION FUNCTION.

No.	**Transformation Types**	**Transformation Function**
1	Nearest neighbour	$f_n(x, y) = g(round(x), round(y))$
2	Linear interpolation	$f_n(x, y) = (1-a)(1-b)g(l,k) + a(1-b)g(l+1,k)$ $+ b(1-a)g(l,k+1) + abg(l+1,k+1)$ $l = round(x), \quad a = x - l$ $k = round(y), \quad b = y - k$
3	Bi-cubic interpolation	$f_n(x, y) = \sum_{l=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} g(l\Delta x, k\Delta y)h_n(x - l\Delta x, y - k\Delta y)$ $h_n = \begin{cases} 1-2\lvert x\rvert^2+\lvert x\rvert^3 & for\ 0\le\lvert x\rvert<1 \\ 4-8\lvert x\rvert+5\lvert x\rvert^2-\lvert x\rvert^3 & for\ 1\le\lvert x\rvert<2 \\ 0 & otherwise \end{cases}$ where h_n is the interpolation kernel and g(.,.) is the sampled version of input image.

III. ISSUES IN BIOMETRIC TRANSFORMATION INVARIANT PATTERNS

The issues in the biometric transformation invariant patterns are such as subsist detection during image acquisition, image quality measure and tilted Biometric patterns have been addressed in the work to achieve rotation-invariant recognition system.

A. Subsist Detection

Subsist detection is required to determine, if the biometric sample is actually presented by living human or from any other artificial sources or not. One of the vulnerable points in the system is the user data capture interface that should ensure the signals for the genuine subject and should contradict artificial sources such as printed picture of biometrics, spurious fingers or eyes, video clips and any kind of objects like eyes. A challenge-response test ensures the pupil diameter variations in the imaging. It monitors the diameter of eye images under diverse lighting conditions that enables the system to prohibit artificial sources.

B. Image Quality Assessment

Eye images can be acquired in widely varied luster under dark or bright lightings or in twilight or sunlight environments. Moreover, eye images might be truncated due to pan/tilt of subjects' head movements, closed eyelashes/eyelids may be portrayed on the Biometric portion, spectacle reflection may occur in the Biometric area and glare images may be acquired due to reflection of fortifications colors. To resolve these problems, image quality must be assessed.

C. Distance Variation in Capturing

The current research includes identification of persons by their Biometrics even as they are walking around the place. Currently, Biometric recognition technologies identify a person who stands in front of a scanner and shows his/her eye properly. This is because Biometric pigments are not explicitly sharp enough to be scanned by the passive scanners in non-invasive manner. However, Biometric technology is more reliable than face recognition but it requires cooperation of subjects who will stand in front of the scanner and line up his/her eye properly. Thus, this paper analyzed Biometric patterns' variations by varying its capturing distances. Remote Biometric recognition (RIR) is a value added solution to the IRS for the public if it is capable of recognizing the persons by their Biometric while moving at a distance. However it requires a high-resolution camera to capture the subjects' eye images without their full cooperation. Thus this paper aims to suggest some technical contribution related to the Remote Biometric recognition system (RIRS).

D. Orientation-invariant Patterns

Most traditional approaches used a set of predefined rotation parameters to match the Biometric scores. For example, -9, -6, -3,0,3,6 and 9 degrees were employed by Li ma et al [1], Li ma et al [2] and seven left and right shifts were carried out by Daugman [3]. But during the runtime these predefined degrees may not be enough to estimate the rotation angle of Biometric patterns and hence produce false positives while pan/tilt angles of head positioning are rapidly varied. This paper estimates the rotation angle of Biometric patterns during imaging. By using the estimation angle, pattern's rotation is corrected to its principal direction and then applies the feature extraction to encode the Biometric features. Thus, transformation invariant Biometric pattern recognition has been achieved efficiently by the system.

IV. TRANSFORMATION INVARIANT BIOMETRIC PATTERN ANALYSIS

In this work, a Biometric camera is used to capture user's eye images. It acquires images by passing Near Infrared (NIR) waves. The acquisition distance is normally between 19 and 36 inches and average capturing time is 1.5 seconds. Due to distance or luster changes, recognition process may produce different recognition rates. Perhaps, the recognition rate of the same candidate's Biometric features may slightly vary at sunlight and twilight criterion. The image acquisition phase should consider three main aspects namely, the lighting system, the positioning system and the physical capturing system. Usually, in enrollment phase the Biometric images are captured without any eyewear that aids to encode the Biometric features accurately. However, the use of eyewear such as spectacles or contact lens, during the verification does not affect the recognition process.

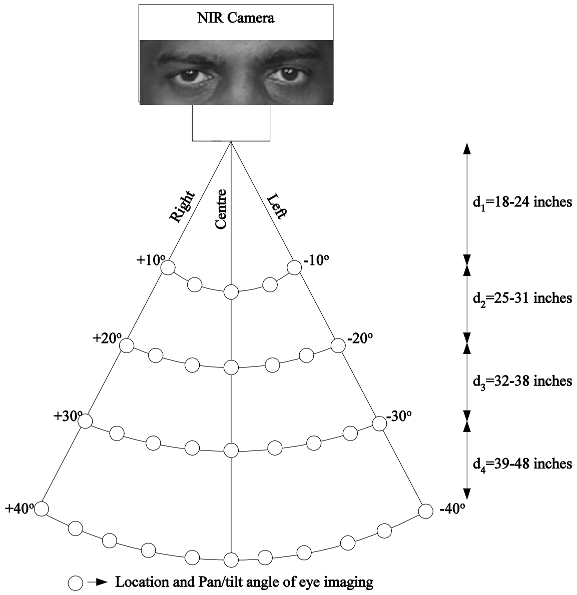


Figure 3. Position of eye imaging at different distance and pan/tilt angles.

Biometric recognition system can work both in outdoor and indoor conditions without any hot spot of lighting intensities. But unlike face, palm and fingerprints imaging, Biometric is an internal organ, which is present inside the closed area of eyelids. Hence the users must provide full cooperation to acquire their eye images. Eye images are acquired in various pan and tilt angles such as, pan ranges from +40 to –40 degrees, tilt varies from 0 to 20 degrees and the distance of imaging varies between d_1=18-24, d_2=25-31, d_3=32-38, and d_4 =39-48 inches. This is illustrated in Fig. 3.

A. Anti-spoofing

Since biometric features may be counterfeited and criminally used to cheat the secured system, Challenge-Response Triangularization-Test (CRTT) is carried out to ensure that images are acquired from actual human beings and not from artificial sources like Biometric photographs, spurious eyes or other artificial sources. Daugman et al. [3], and Li ma et al [1] [2] had discussed about these issues, but they did not present any specific scheme. Challenge-Response-Trigularization Test (CRTT) is suggested to improve anti-spoofing.

The human eye is controlled by two muscles, namely, the dilator and sphincter that allow eye to adjust its size and controls the amount of light entering the eye. The eye acts like a shutter of the camera. The pupil is normally a circular hole in the middle of the iris. Like a camera aperture, Biometric regulates the amount of light passing through the retina at the backside of the eye. If the amount of light entering the eye is increased, for example bright light, the iris sphincter muscle pulls towards the center; the size of the pupil is diminished and allows less light to reach the retina. If the amount of light entering the eye is decreased such as in the dark or at night, then iris dilator muscle pulls away from the center, the size of the pupil is expanded and allows more light to reach the retina. CRT uses these biological facts and verifies the response of the pupil diameter by varying luster levels from the same distance of capturing. Fig. 4 shows computation of the pupil diameter during acquisition.

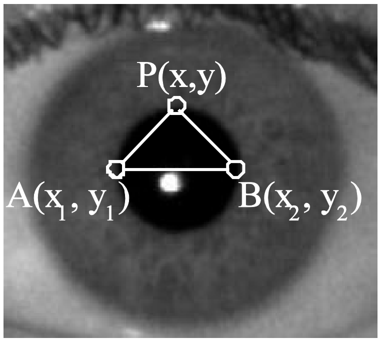


Figure 4. Computation of pupil area dissimilarity due to luster variations based Challenge-Response-Triangularization-Test.

Following is an algorithm for computation of pupil diameter:

Step 1: Acquire eye image in three diverse lighting levels from the same standing distance of the subject.

Step 2: Compute Triangularization of the pupil diameters. If diameters are divergent then let the image is in point of fact sourced from real-life scenario otherwise artificial sources may be counterfeited and alarm for catastrophe. The diameter of the pupil (Φ) is calculated by (8) – (10).

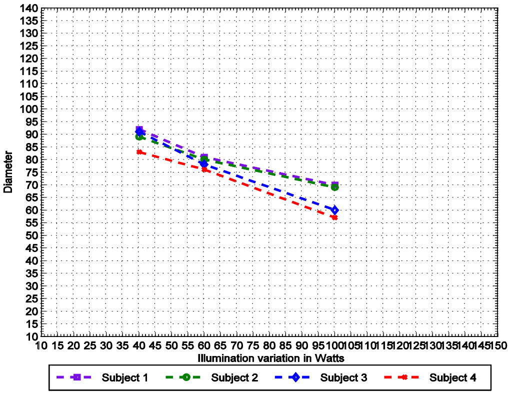
$$\Phi = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \qquad (8)$$

$$T_d = \sum_{i=0}^{n-1} |\Phi_i - \Phi_{i+1}| \leq \kappa, \qquad (9)$$

$$\partial = \begin{cases} 1 & T_d \neq 0 \,\&\, T_d \neq \infty \\ 0 & Otherwise \end{cases}, \qquad (10)$$

where T_d is a total diameter difference of the pupil in capturing n sequences, ∂ is a Challenge-Response Trigularization Test variable, n is the number of eye images, Φ_i and Φ_{i+1} are diameters of the pupil in different luster. The CRTT identifies the input from a real-life scenario if ∂ = 1, otherwise it decides that input is coming from artificial sources.

In an involuntary response test such as CRTT, the user's body automatically provides the response with physiological changes or reaction to a stimulus. For this test, 10 printed photos were directly captured by the Biometric camera and used as input. No changes were observed in pupil diameters. Thus, the system found that artificial sources were counterfeited.100 subjects' eye images were captured with different luster of 40 watts, 60 watts and 100 watts light sources. The pupil diameter of 'subject1' was observed to be 83, 75 and 69 pixels with luster from 40 watts, 60 watts and 100 watts light sources respectively. The result show that pupil diameters vary from 92 to 82 pixels in 40 watts, from 81 to 69 pixels in 60 watts and from 70 to 55 pixels in 100 watts luster. It is noted that luster and pupil diameter sizes are inversely proportional. The result of diameter variations of luster and pupil diameter variation are shown in Figs. 5a and 5b respectively.

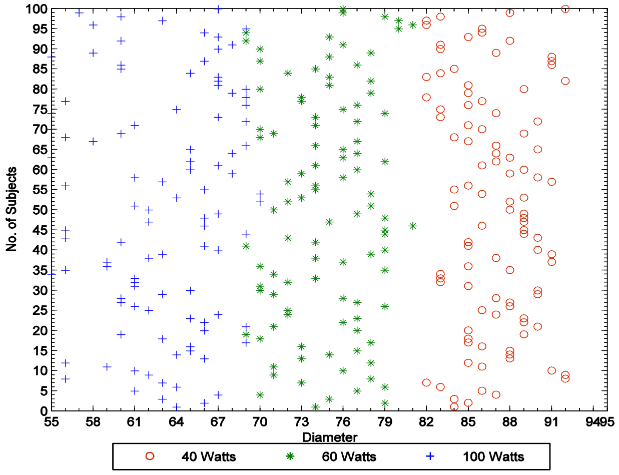


(a)

Figure 5a. Results of Challenge-Response-Triangularization-Test Variation of diameter with diverse luster.

B. *Image prominence checking*

In real-life acquisition, images are acquired in different kinds of distances such as eyes are captured from 18 to 48 inches in the non-invasive mode images are acquired from 2 to 20 feet. Hence, these images are exaggerated by various artefacts such as motion blurred, defocused, truncation of AROI and other luster issues. These artefacts produce misclassification in the recognition phase. If an image can exceed a minimum focus threshold then it will be used for further processing. The image prominence-checking module is discussed to choose a best frame for negating the over truncated, spectacle reflected and closed eyelashes/eyelid images. The blur correction process is performed to get back non-blurred images.



(b)

Figure 5b. Results of CRTT Pupil diameter variation. Red, Green and Blue circles indicate 40, 60 and 100 watts variations, respectively.

C. Biometric image status checking

This phase is invoked especially for status checking of eye images while acquiring in non-invasive mode. There is no direct contact between the Biometric acquisition camera and candidate Biometric. In this mode, there are many possibilities to acquire closed eyelashes/eyelids images or defocused images, truncation of Biometric images, spectacle-contact lens reflection images and glare images. Hence, image status checking method helps to choose a moderate image for further processing. The following algorithms are used to select a moderate image from the acquisition. Finding closed eyelashes/eyelids images: In the closed eye images, Biometrics pattern is not focused properly. For that the system checks the fraction of closed portion of the Biometric. Thus it calculates gray magnitude of the eye area. If the up ceiling average of gray magnitude of pupil is less than the threshold (η_a) then the given image has closed eyelashes/eyelids otherwise not occluded. Equation (11) describes the process.

$$\sum_i^n\sum_j^m\left\langle\frac{\theta(i,j)}{\psi_{nm}}\right\rangle\le\lambda_a, \qquad (11)$$

where n and m is the size extracted portion of eye image, λ_a is an adaptive threshold value of pupil area and $\theta(i,j)$ is an array of eye image gray magnitudes.

Truncation of Biometric: During image acquisition, the Biometric portion may be truncated due to the alignment of head positions. These types of images may produce false alarm in the Biometric matching processes. Hence the system verifies the threshold limit of truncation of the Biometric portion, which is tolerable to the feature extraction and classification phases. First a rectangle outline is fixed on the eye image and the scanning process begins to pass through the perimeter of a rectangle. While travelling, count the number of magnitudes that are satisfied with the Biometric threshold values. If the number of pixels is above the tolerance level i.e., truncation is present, then image is rejected, otherwise the image is accepted for further processing.

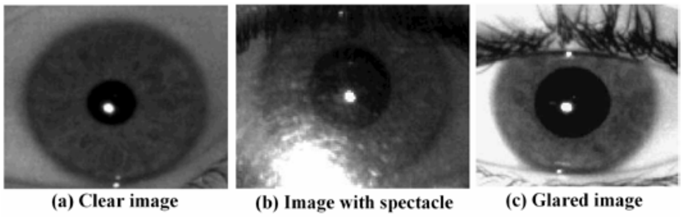
Spectacle reflection and glare eye images: Spectacle reflection and stare angrily position may appear in Biometric area. These types of artifacts occur due to over lighting conditions or sunlight reflection. Sometimes, spectacle can glance off large amount of lighting intensity that may reflect on Biometric portion. These issues may produce false positives in the recognition phase and the problems are overcome by filtering glare and spectacle reflection. Figure 5 shows some possible artifacts of images. Figure 7 depicts histogram corresponding to images in Fig. 6.

$$d_i=\sqrt{(X_{i+1}-X_i)^2+(Y_{i+1}-Y_i)^2}\quad i=1,\ 2,\ 3,\ 4\ (\mathrm{mod}\ \ 4), \qquad (12)$$

$$\left[\sum_{i=1}^{4}(\sum_{j=1}^{d_i}\theta(i,j)\le\lambda)\right]\ge\xi, \qquad (13)$$

where d_i is the distance between end points of ith line segment, λ is the Biometric threshold value and ξ is the tolerable level of the image.

Light incident on a flat surface will be reflected and transmitted. When a camera captures opaque non-luminous objects, the total light reflected is the sum of the contributions from the light sources and other reflecting surfaces in the frame. Often, light sources can be light-emitting sources and reflecting sources as the walls of a room. A luminous object reflection depends on both light source and light reflector. For example spectacle surface that are rough or grainy, reflects light in all directions. This scattered light is called diffuse reflection. A very rough matte surface produces predominantly diffuse reflections, so that the surface appears equally bright from all viewing directions.



(a) Clear image (b) Image with spectacle (c) Glared image

Figure 6. Diverse Luster occurs in the real-life biometric patterns while acquisition.

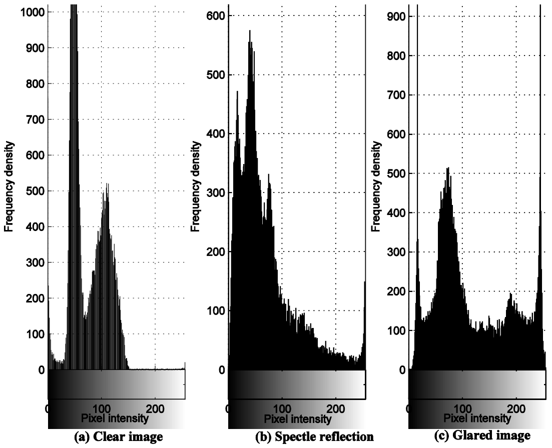
In addition to diffuse reflection, light sources create, bright spots called specula reflection. These bright spots occur on shiny surfaces than on dull surfaces. Therefore, this system finds the adaptive local and global luminance level of the captured eye images followed by checking the tolerable level of the images.

To achieve consequently, first we extract a rectangle portion from the image eye that contains Biometric part. Next, convert this area into binary image and compute its arithmetic mean values. If mean values of reflection of a glass or glare are within the tolerance level, then the image is used for further processing, otherwise the image is rejected. This is described in (14) – (15).

$$R(i, j) = \begin{cases} 0 \ \ if \ \theta(i, j) \leq l_u (Background) \\ 1 \quad \ \ otherwise(Foreground)\text{'} \end{cases} \tag{14}$$

$$\left[\frac{1}{\psi_{nm}} \sum_{i=1}^{N} \sum_{j=1}^{M} R(i, j) \right] \leq \mu_t , \tag{15}$$

where n and m is size of the rectangle, $R(i, j)$ is a set of binary values in the bounded rectangle, l_u is a luster tolerance level and μ_t is the mean threshold value. After checking the captured images, this phase chooses a portion of the image, which is less than or equal to the threshold value for localization process.



(a) Clear image

(b) Spectle reflection

(c) Glared image

Figure 7. Histogram of the Clear (a), Spectacle reflection (b) and Glared (C).

D. Luminance level analysis

Intensity level of real-life images is not having unique thresholds for pre-processing. Hence luster level analysis is required to choose the intensity levels of the image while acquiring from different light sources. This approach incorporates both optimal and adaptive thresholds that aid to perfectly choose the threshold values for the binarization and localization processes. In this paper, normal distribution based optimal threshold, local characteristics based local and global threshold analyses have been carried out to select the best thresholds for the real-life images. In addition, fusion of mean intensity analysis is calculated to adopt thresholds in a widely anecdotal luster.

V. Conclusion

This research paper characterizes diverse ways to capture biometric images in real-life scenarios, which sustain to get potent recital in the transformation invariant pattern analysis. Locating AROP is accomplished in the diverse state of affairs. These investigations unshackle an innovative tactic of research in terms of biometric recognitions such as in face and iris. The anti-spoofing was done by the unintentional response of physiological changes or reaction to a stimulus of the body, which helps to prevent artificial sources, enrolling or verifying by the system. The image pose checking was used to bear out the scenery of the eye images in diverse luster and artifacts. Finally, the luster levels of the images have been estimated. In further research, a global and local optimal doorsill will be suggested in incarnation with mean analysis. This method will be exploited to speculate doorsill values for localization or binarization process of the system.

Acknowledgment

Authors thank their family members and children for their continuous support and consent encouragement to do this research work successfully.

References

[1] Li ma, Tieniu Tan, Yunhong Wang and Dexin Zhang, "Personal Identification Based on Biometric Texture Analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 12, pp. 1519-1533, 2003.

[2] Li ma, Tieniu Tan Yunhong Wang and Dexin Zhang, "Efficient Biometric Recognition by Characterizing key Local variations," IEEE Transaction on Image Processing, Vol. 13, No. 6, pp. 739-750, 2004.

[3] Daugman J., "How Biometric Recognition Works," IEEE Transactions On Circuits and Systems For Video Technology, Vol. 14, No. 1, pp. 21-30, 2004.

[4] Bremananth R. and Chitra A, "A new approach for Biometric pattern analysis based on wavelet and HNN," Journal of Computer Society of India, Vol. 36, No.2, pp. 33-41(ISSN: 0254-7813), 2006.

[5] Bremananth R. and Chitra A, “Rotation Invariant Recognition of Biometric,” Journal of Systems Science and Engineering, Systems Society of India, Vol.17, No.1, pp.69-78, 2008.

[6] Bremananth R, Ph.D. Dissertation, Anna University-Chennai, PSG College of Technology, India, 2008.

[7] Milan Sonka, Vaclav Hlavac and Roger Boyle, Image processing, analysis, and Machine Vision, Second edition, ITP Press, USA, 1999.

AUTHORS PROFILE

Bremananth R received the B.Sc and M.Sc. degrees in Computer Science from Madurai Kamaraj and Bharathidsan University in 1991 and 1993, respectively. He obtained M.Phil. Degree in Computer Science and Engineering from GCT, Bharathiar University, in 2002. He received his Ph.D. degree in 2008 from Department of Computer Science and Engineering, PSG College of Technology, Anna University, Chennai, India. He has completed his Post-doctoral Research (PDF) from the School of Electrical and Electronic Engineering, Information Engineering (Div.) at Nanyang Technological University (NTU), Singapore, 2011. Before joining NTU, Singapore, he was a Professor and Head, Department of Computer Science and Application, Sri Ramakrishna Engineering College, in India. He has 18+ years of experience in teaching, research and software development at various Institutions. Currently, He is an Assistant Professor of Information Technology department, Sur University College, Sur, Oman, affiliated to Bond University Australia. He received the M N Saha Memorial award for the best application oriented paper in the year 2006 by Institute of Electronics and Telecommunication Engineers (IETE). His continuous contribution of research was recognized by Who’s who in the world, USA and his biography was published in the year 2006. He is an associate editor of various International Journals in USA and He is an active reviewer of various IEEE International conferences/Journals. His fields of research are Acoustic holography, Acoustic imaging, Biometrics, Computer Vision, Computer network, Image processing, Microprocessors, Multimedia, OCR, Pattern recognition, Soft Computing and Software engineering.

Dr. Bremananth is a member of Indian society of Technical Education (ISTE), Advanced Computing Society (ACS), International Association of Computer Science and Information Technology (IACIT) and Institute of Electrical and Telecommunication Engineers (IETE). He can be reached at bremresearch@gmail.com.

Improving the Quality of Applying eXtreme Programming (XP) Approach

Nagy Ramadan Darwish
Department of Computer and Information Sciences,
Institute of Statistical Studies and Research,
Cairo University, Cairo, Egypt
drnagyd@yahoo.com

Abstract—This paper is focused on improving the quality of applying eXtreme Programming (XP) approach on software development process. It clarifies the fundamentals of agile methods of software development. It presents the basic concepts and features of XP approach. XP approach can be viewed as life cycle phases that include six phases: exploration, planning, iterations to release, production, maintenance, and death. Each XP phase can be achieved through performing a set of steps. In this paper, the researcher develops a set of elaborated steps for achieving each XP phase. In addition, the researcher proposes a quality assurance approach for applying XP approach. The proposed quality assurance approach can be used for assuring the quality of achieving XP phases. Then, the deviation between the actual quality and the acceptable quality level can be identified and analyzed. The weaknesses of the software development practices can be discovered, treated to improve the quality of each phase, and avoided in further phases. The strengths of the practices can be discovered, utilized, and encouraged.

Keywords- *eXtreme Programming; XP Approach; Agile Methods; Software Development; Quality Evaluation; Improvements*

I. Introduction and Problem Definition

Software development is a mentally complicated task. Therefore, different software development methodologies and quality assurance methods are used in order to attain high quality, reliable, and bug free software [17]. In recent years, agile software development methods have gained much attention in the field of software engineering [27]. A software development method is said to be an agile software development method when a method is people focused, communications-oriented, flexible (ready to adapt to expected or unexpected change at any time), speedy (encourages rapid and iterative development of the product in small releases), lean (focuses on shortening timeframe and cost and on improved quality), responsive (reacts appropriately to expected and unexpected changes), and learning (focuses on improvement during and after product development) [1].

Agile software development is an iterative and incremental approach that is performed in a highly collaborative manner to produce high quality software that meets the changing needs of its stakeholders. Agile software development methods offer a viable solution when the software to be developed has fuzzy or changing requirements, being able to cope with changing requirements throughout the life cycle of a project [7]. Agile software development methods include XP, Scrum, Crystal, Feature Driven Development (FDD), Dynamic System Development Methodology (DSDM), and Adaptive Software Development (ASD) [4].

- XP is the best known agile method that is driven by a set of shared values including simplicity, communication, feedback and courage. The XP values, practices, and life cycle will be explained in the next section of this paper.
- Scrum is an iterative and incremental approach for managing the software projects in a changing environment. Each iteration aims to produce a potential set of the software functionality.
- Crystal methodologies focus on incremental development which may be in parallel. Each increment may take several iterations to complete. The tunable project life cycle that is common for all Crystal methodologies is: envisioning, proposal, sales, setup, requirements, design and code, test, deploy, train, alter [1]. Crystal family of methodologies provides guidelines of policy standards, tools, work project, and standards and roles to be followed in the development process.
- FDD is a model-driven and short-iteration approach for developing software. It focuses on the design and building phases. FDD provides guidelines, tasks, techniques and five sequential processes: Develop an Overall Model, Build a Feature List, Plan by Feature, Design by Feature and Build by Feature [24].
- DSDM provides a framework that supports rapid, iterative and collaborative software development for producing high quality business information systems solutions [15]. The basic principle of DSDM is that the resources and timeframe are adjusted and then the goals and the required functionality are adjusted accordingly.
- ASD offers an agile and adaptive approach to high-speed and high-change software projects. ASD replace the static plan-design life cycle by a dynamic speculate-collaborate-learn life cycle. ASD focuses more on results and their quality than the tasks [13].

XP is one of the most popular agile development methods. Therefore, it is the main concern of this paper. The XP process is characterized by short development cycles, incremental planning, continuous feedback, and reliance on communication and evolutionary design [27]. It is designed for use with small teams that need to develop software quickly and in an environment of rapidly changing requirements.

Although the many advantages and features of XP approach, using it for developing software doesn't guarantee the success of this process at an acceptable level of quality. In addition, software projects are faced with many challenges that may lead them to the failure. Therefore, there is a need for assuring the quality of software development. Quality assurance is all the planned and systematic activities implemented within the quality system, and demonstrated as needed, to provide adequate confidence that an entity will fulfill the requirements for quality [11]. This paper focuses on elaborating a set of steps for achieving each XP phase, evaluating the quality of achieving this phase, and determining the deviation of achieving the phase to improve the quality of software development.

II. eXtreme Programming (XP) Approach

Extreme Programming was developed at Chrysler by Kent Beck while working on a payroll project as a member of a 15 person team. Beck continued to refine and improve the XP methodology after the project was completed until it gained worldwide acceptance in 2000 and 2001 [14].

The XP software development process focuses on iterative and rapid development. XP approach stresses communication and coordination among the team members at all times; and requires cooperation between the customer, management and development team to form the supportive business culture for the successful implementation of XP [1]. It is designed for use in an environment of rapidly changing requirements. It helps to reduce the cost of change by being more flexible to changes. XP is characterized by six phases: exploration, planning, iterations to first release, productionizing, maintenance and death. XP is a software development discipline in the family of agile methodologies that contributes towards quality improvement using dozen practices [17]. XP consists of twelve practices, which are planning game, small releases, metaphor, simple design, testing, refactoring, pair programming, collective code ownership, continuous integration, 40-hour week, on-site customer, and coding standard [19]. Figure (1) illustrates the XP values, practices, and phases.

A. *XP Values*

XP is driven by a set of values including simplicity, communication, feedback and courage.

- Communication: An Agile method emphasises on face-to-face communication within the team and with the customer who is closely involve with the development process [21]. XP requires direct communication among all members to give the developers a shared view of the system which matches the view held by the users of the system.
- Feedback: The software developers should always have a way for getting information about the development process. Feedback relates to many dimensions that include the system, customer, and team. Feedback from the system and the team aims to provide project leaders with quick indicators of the project's progress to take corrective or supportive actions. In addition, feedback from customer includes the functional and acceptance tests.

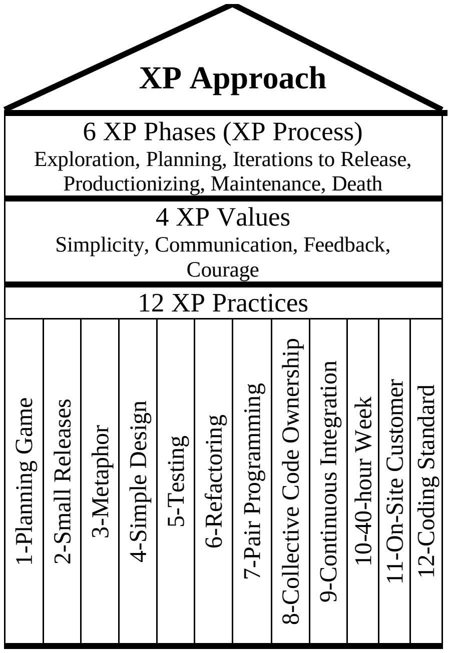


Figure (1): XP Values, Practices, and Phases.

- Simplicity: A simple design always takes less time to finish than a complex one. Therefore, XP encourages starting with the simplest solution. Extra functionality can then be added later. Extreme programmers do the simplest thing that could possibly work, and leave the system in the simplest condition possible. This improves the overall speed of development while still retaining an emphasis on working software.
- Courage: Courage means that developers are prepared to make important decisions that support XP practices. Courage enables developers to feel comfortable with refactoring their code when necessary. This means reviewing the existing system and modifying it so that future changes can be implemented more easily. In addition, courage may include removing source code that is obsolete, no matter how much effort was used to create that source code.

B. *XP Practices (rules)*

The four core values of XP are implemented with twelve core practices: Planning Game, Small Releases, Metaphor, Simple Design, Testing, Refactoring, Pair Programming, Collective Code Ownership, Continuous Integration, 40-hour Week, On-Site Customer, and Coding Standard.

1. Planning Game: At the beginning of the development process, customers, managers, and developers meet to create, estimate, and prioritize requirements for the next release. The requirements are captured on “story cards” in a language understandable by all parties. In fact, the developers estimate the effort needed for the implementation of customers’ stories and the customers then decide about the scope and timing of releases. The planning game and the story cards offer the devices to perform planning on the most detailed level for very short periods of time [18].
2. Small Releases: The development is divided in a sequence of small iterations, each implementing new features separately testable by the customer [7]. XP increases the pace of the delivery of the software by having short releases of 3-4 weeks. At the end of each release, the customer reviews the software product, identify defects, and adjust future requirements. An initial version of the software is put into production after the first few iterations. The small releases help the customer to gain confidence in the progress of the project. In addition, the small releases help the customer to come up with their suggestions on the project based on real experience.
3. Metaphor: The system metaphor is the story that customers, developers, and managers can tell about how the system works [19]. The system metaphor is an effective way of getting all members of the project team to visualize the project. It should provide inspiration, suggest a vocabulary, and a basic architecture. This is the only principle not strictly required in every XP project.
4. Simple Design: The developers must focus on designing only what is needed to support the functionality being implemented. The Developers are urged to keep design as simple as possible, say everything once and only once. A program built with XP should be a simple program that meets the current requirements. Kent Beck stated that the right design for the software at any given time is the one that runs all the tests, has no duplicated logic, states every intention important to the programmers, and has the fewest possible classes and methods [19].
5. Testing: Testing is an integral part of XP. All code must have automated unit tests and acceptance tests, and must pass all tests before it can be released [7]. The tests are written before coding. Sometimes, this practice is called “test first”. Programmers write unit tests so that their confidence in the operation of the program can become part of the program itself. For the same reason, customers write functional tests. The result is a program that becomes more and more confident over time.
6. Refactoring: Refactoring is the process of changing the code in order to improve it by removing redundancy, eliminating unused functionalities, improving code readability, reducing complexity, improving maintainability, adapting it to patterns or even trying to make the software work in an acceptable way. Refactoring throughout the entire project life-cycle saves time of development and increases quality.
7. Pair Programming: Pair programming is one of the key practices of XP. It is a programming technique that requires two programmers to work together at solving a development task while sharing the monitor, the keyboard, and the mouse. The work may include analyzing data, creating the data model, programming, etc. The advantages of pair programming are improving productivity, the quality of the solution, and job satisfaction [26]. Moreover, it reduces the time needed for task completion, it is particularly useful in complex tasks, and it is useful for training.
8. Collective Code Ownership: This practice indicates that the code is owned and shared by all developers. Everyone is able to edit it and see the changes made by others. It tends to spread knowledge of the system around the team. The code should be subjected to configuration management.
9. Continuous Integration: Developers integrate a new piece of code into the system as soon as possible it is ready. All tests are run and they have to be passed for accepting the changes in the code. Thus, XP teams integrate and build the software system multiple times per day. Continuous integration reduces development conflicts and helps to create a natural end to the development process.
10. 40-Hour Weeks: This practice indicates that the software developers should not work more than 40 hour weeks, and if there is overtime one week, that the next week should not include more overtime. The people perform best and most creatively if they are rested, fresh, and healthy. Therefore, requirements should be selected for iteration such that developers do not need to put in overtime.
11. On-Site Customer: A customer works with the development team at all times to answer questions, perform acceptance tests, and ensure that development is progressing as expected. This customer-driven software development led to a deep redefinition of the structure and features of the system [7]. It supports customer-developer communication [18].
12. Coding Standards: This practice indicates that the developers must agree on a common set of rules enforcing how the system shall be coded. This makes the understanding easier and helps on producing a consistent code. Coding standards are almost unavoidable in XP, due to the continuous integration and collective ownership properties.

C. XP Process

XP approach can be viewed as life cycle phases that include six phases: exploration, planning, iterations to release, productionizing, maintenance, and death [1]. Each phase can be achieved through a set of activities. Figure (2) illustrates the XP life cycle [22].

1. Exploration Phase: In the exploration phase, the customers write out the story cards that they wish to be included in the first release. Each story card describes a feature to be added into the program. At the same time, the development team gets familiar with the development

environment and the addressed technology [25]. The exploration phase takes between a few weeks to a few months, depending largely on how familiar the technology is to the programmers.

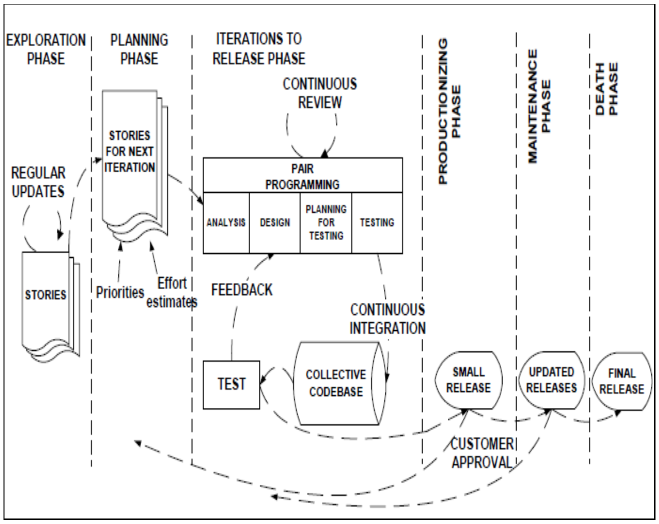


Figure (2): XP Life Cycle [22].

2. Planning Phase: In the planning phase, the customers set the priority order for the stories and an agreement of the features of the first small release is made. The developers estimate the necessary effort and time for each story. Then the schedule of the first release is developed and approved. The planning phase takes a couple of days.
3. Iterations to Release Phase: In the iterations to release phase, the actual implementation is done. This phase includes several iterations of the systems before the first release. The schedule is broken down to a number of iterations that will each take one to four weeks to implement [22]. For each iteration, the customer chooses the smallest set of most valuable stories that make sense together and programmers produce the functionality. Small releases reduce the risk of misled development. XP coding always begins with the development of unit tests. After the tests are written, the code is developed and continuously integrated and tested. At the end of the iteration all functional tests should be running before the team can continue with the next iteration [19]. When all iterations scheduled for a release are completed the system is ready for production.
4. Productionizing phase: The production phase includes extra testing and checking of the functionality and performance of the system before the system can be released to the customer [22, 25]. At this phase, new changes may still be found and the decision has to be made if they are included in the current release. During this phase, the iterations may need to be quickened from three weeks to one week. The postponed ideas and suggestions are documented for later implementation during, e.g., the maintenance phase.
5. Maintenance Phase: After the first release is productionized, the system must be kept running in production, while remaining stories are implemented in further iterations. Therefore, the maintenance phase requires an effort for customer support tasks. Development stays in this phase until the system satisfies the customers' needs in all aspects.
6. Death Phase: Finally, development enters the death phase when the customer has no more stories to be implemented, and all the necessary documentation of the system is written as no more changes to the architecture, design, or code are made. Death may also occur if the system is not delivering the desired outcomes, or if it becomes too expensive for further development.

III. The Elaborated Steps for Achieving XP Phases

In XP approach, developers communicate among each other to efficiently utilize tacit knowledge and quickly find new solutions to current challenges. Developers communicate with customer representatives to deliver the most valued features, gain rapid feedback on deliveries and improve the customer's trust and confidence [23].

XP approach can be viewed as life cycle phases that include six phases: exploration, planning, iterations to release, productionizing, maintenance, and death [1]. Each phase can be achieved through a set of steps. The researcher elaborates a set of steps for achieving each phase. In this section, the elaborated steps are presented. In the elaborated steps, if we don't tell who is responsible for performing the step, we mean that the developers and customers together must participate in doing it.

A. The Elaborated Steps of "The Exploration Phase"

The XP software development process is regarded as the flow in which user stories are generated, designed, coded and unit tested, refactored and verified. A user story is a software system requirement formulated as one or two sentences in the everyday or business language of the customers. The user stories should be written by the customers for a software project. During the development process, customers can generate new user stories and change old ones [27]. The elaborated steps required for achieving the exploration phase are:

1. Presenting and clarifying the purpose and the steps of "the exploration phase" to the customers who participating in the team.
2. Obtaining a preliminary background of the project. The background will be incremented through the next phases. The project's background includes project's motivation, assumptions, constraints, addressed technology, and the acceptance criteria.
3. Clarifying the purpose of the story cards as a tool for collecting the requirements. Each story card describes a feature to be added into the current release.
4. Presenting and clarifying the writing standards that must be considered when writing the story cards. For example, the

stories must be consistent, clear, testable, and integrated with the other related stories.

5. Writing the story cards that the customers wish to be included in the current release. This step must be done by the customers.
6. Understanding the story cards. This step must be done by the developers.
7. Analyzing and validating the story cards.

B. The Elaborated Steps of "the Planning Phase"

In the planning phase customers assign priorities to their stories and developers estimate the necessary effort for their implementation. Then a set of stories for the first small release is agreed upon and the release is scheduled according to the programmers' estimations [25]. If possible, near-site customers should do this with programmers in face-to-face meetings [20]. The elaborated steps required for achieving the planning phase are:

1. Presenting and clarifying the purpose and the steps of "the planning phase" to the customers who participating in the team.
2. Setting the priority order of the stories. This step must be done by the customers.
3. Identifying and negotiating the features that must be included in the current release.
4. Preparing an approved list of features needed to implement the current release.
5. Estimating the necessary effort and time for each story.
6. Preparing a proposed schedule for the current release.
7. Negotiating and approving the proposed schedule of the first release to reach to a final one.

C. The Elaborated Steps of "Iterations to Release Phase"

XP promotes the concept of "small releases" [16]. The meaningful releases should be made available to users when completed. This will allow early and frequent feedback from the customers. The elaborated steps required for achieving this are:

1. Presenting and clarifying the purpose and the steps of "iteration to release phase" to the customers who participating in the team.
2. Breaking down the schedule to a number of iterations. The iteration will take one to four weeks.
3. Choosing the smallest set of most valuable stories that make sense together [25] and useful to be included in each iteration.
4. Reviewing the functionality of all iterations.
5. Selecting the iteration to be implemented. The selection process depends on the logical sequence of the current release's functionalities.
6. Developing the unit tests for the selected iteration.
7. Writing the code for the selected iteration.
8. Integrating and testing the selected iteration.
9. Ensuring that all functional tests were done before moving to the next iteration.
10. Ensuring that all iterations scheduled are completed.
11. Delivering the current release to production phase.

D. The Elaborated Steps of "Productionizing Phase"

In productionizing phase, there are more testing and checking of the functionality and performance of the system such as system testing, load testing, and installation testing. The elaborated steps required for achieving productionizing phase are:

1. Presenting and clarifying the purpose and the steps of "the productionizing phase" to the customers who participating in the team.
2. Performing extra testing and checking of the functionality and performance of the system such as system testing, load testing, and installation testing.
3. Identifying new changes needed to be included in the current release.
4. Implementing and testing the new changes identified in the previous step.
5. Identifying and documenting the postponed ideas and suggestions to implement them during maintenance phase or in next releases.
6. Delivering the current running release to the customers.

E. The Elaborated Steps of "Maintenance Phase"

During the maintenance phase the system must be kept running in production, while remaining stories are implemented in further iterations. Development stays in this phase until the system satisfies the customers' needs in all aspects [25]. The maintenance efforts can be viewed in five main activities: system maintenance, solving system crash, end-user assistance, system enhancement, and system reengineering. The elaborated steps required for achieving maintenance phase are:

1. Presenting and clarifying the purpose and the steps of "the maintenance phase" to the customers who participating in the team.
2. Identifying, analyzing, and documenting the circumstances that led to bugs and symptoms of the problems. Then edit programs to fix bugs.
3. Performing unit, system, and regression testing for the edited programs.
4. Identifying, analyzing, and documenting the causes of the system crash.
5. Identifying and clarifying corrective instructions that are required to prevent the system crash. These instructions may include: terminate the on-line session, reinitialize the application, recover lost or corrupted databases, fix problems of local or wide network, and/or fix hardware problems.
6. Providing users with additional training.
7. Identifying and documenting enhancement ideas and requests.
8. Taking decisions about the enhancement ideas and requests that must be implemented in this phase or moved to next releases.
9. Writing and testing code for the approved enhancement ideas and requests.

F. The Elaborated Steps of "Death Phase"

In the death phase, the software development process has been finished. Now there is no change to architecture, design or code will be made. The elaborated steps required for achieving death phase are:

1. Presenting and clarifying the purpose and the steps of "the maintenance phase" to the customers who participating in the team.
2. Ensuring that all predefined stories has been implemented.
3. Finalizing all project documentation.
4. Evaluating the quality of the current release and the related parts of the system.
5. Identifying and documenting the learned lessons from the project.
6. Studying the feasibility of continuing the running of the release and the system.

IV. The Proposed Approach for Improving the Quality of Applying XP Approach

Applying XP approach on software development process doesn't guarantee the success of this process at an acceptable level of quality. In addition, software projects are faced with many challenges that may lead them to the failure. Therefore, there is a need for assuring the quality of software development. Quality assurance is all the planned and systematic activities implemented within the quality system, and demonstrated as needed, to provide adequate confidence that an entity will fulfill the requirements for quality [11].

The researcher proposes a quality assurance approach for applying XP approach. The proposed quality assurance approach can be used for assuring the quality of achieving XP phases. Figure (3) illustrates the proposed quality assurance model. The proposed quality assurance approach includes the following activities:

1. Achieving XP phase using the elaborated steps.
2. Evaluating the quality of the achieved phase.
3. Identifying the deviation between the actual quality and the acceptable quality level.
4. Analyzing the deviation to take corrective or supportive actions.

Firstly, the developers must recall, present and clarify the elaborated steps of the current XP phase to the customers participated in the XP team. Then, the developers and customers begin to achieve the current XP phase using the elaborated steps. The elaborated steps of each phase are not having the same level of importance. Each step may have one of the cases: high importance, average importance, or low importance. Secondly, the quality of the achieved phase must be evaluated using the common statistical techniques for measuring the quality. Thirdly, the deviation between the actual quality and the acceptable quality level must be identified. The acceptable quality level differs from project to another depending on the project field and the acceptance criteria of customers. Fourthly, corrective actions must be done to the current phase if the deviation due to a weakness of the performance. Otherwise, supportive actions may be needed for the next phases.

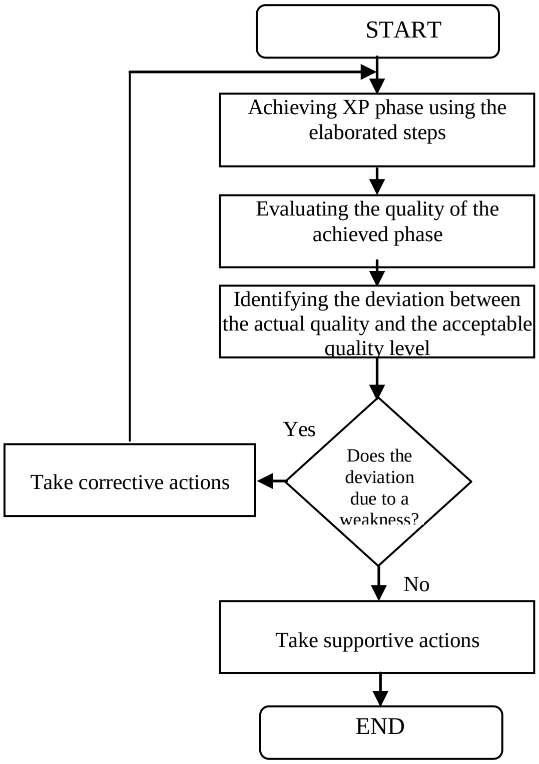

START
Achieving XP phase using the elaborated steps
Evaluating the quality of the achieved phase
Identifying the deviation between the actual quality and the acceptable quality level
Does the deviation due to a weakness?
Yes
Take corrective actions
No
Take supportive actions
END

Figure (3): The Proposed Quality Assurance Approach.

V. Conclusion

The main objective of this paper was improving the quality of applying XP approach. Therefore, the researcher elaborates a set of steps for achieving each XP phase and proposes a quality assurance approach for applying XP approach. The developers and customers can use the elaborated steps as a guiding tool for achieving each XP phase. The proposed quality assurance approach can be used for assuring the quality of achieving each XP phase. Then, the deviation between the actual quality and the acceptable quality level can be identified and analyzed.

We conclude that the quality assurance practices play a very important role for increasing the probability of the software development success. Applying the XP approach for developing software doesn't guarantee the success of this process. Therefore, there is a need for complementary quality assurance practices.

VI. Future Work

There are many efforts can be done in the field of XP approach in the future. Briefly, the following points are expected to be focused:

- Proposing an approach for evaluating the quality of XP phases.

- Building a software tool for managing XP projects.
- Using XP approach to achieve higher levels in Capability Maturity Model Integration (CMMI) for IT companies.
- Enhancing the calculation of software metrics related to XP projects.

REFERENCES

[1] A. Qumer and B. Henderson-Sellers, "An Evaluation of the Degree of Agility in Six Agile Methods and its Applicability for Method Engineering", Information and Software Technology Vol. 50 Issue 4, 2008, pages 280–295, 2008.

[2] Alan S. Koch, "Agile Software Development - Evaluating the Methods for Your Organization", Artech House INC., 2005.

[3] Beck, K. and Andres, C., "Extreme Programming Explained: Embrace Change", Addison-Wesley, 2005.

[4] Dean Liffingwell, "Scaling Software Agility – Best Practices for Large Enterprises", The Agile Software Development Series, Pearson Education Inc., 2007.

[5] G. Gordon Schulmeyer, "Handbook of Software Quality Assurance", 4th edition, Artech House Inc., 2008.

[6] Gary Chin, "Agile Project Management: How to Succeed in the Face of Changing Project Requirements", AMACOM, 2004.

[7] Giulio Concas, Marco Di Francesco, Michele Marchesi, Roberta Quaresima, and Sandro Pinna, "An Agile Development Process and Its Assessment Using Quantitative Object-Oriented Metrics", 9th International Conference, XP 2008, Limerick, Ireland, Proceedings, June 2008.

[8] Hamid Mcheick, "Improving and Survey of Extreme Programming Agile Methodology", International Journal of Advanced Computing (IJAC), Vol. 3, Issue 3, July 2011.

[9] Helen Sharp and Hugh Robinson, "Collaboration and co-ordination in mature eXtreme programming teams", International Journal of Human-Computer Studies 66 pages 506–518, 2008.

[10] Hulkko, H. and Abrahamsson, P., "A Multiple Case Study on the Impact of Pair Programming on Product Quality", Proceedings Of ICSE, pp. 495–504, 2005.

[11] Ince, Darrel, "Software Quality Assurance - a Student Introduction", McGraw-hill international (UK) limited, 1995.

[12] Ioannis G. Stamelos and Panagiotis Sfetsos, "Agile Software Development Quality Assurance", Information science reference, Idea Group Inc., 2007.

[13] James A. Highsmith, "Adaptive Software Development: A Collaborative Approach to Managing Complex Systems", Dorset House Publishing, New York, 2000.

[14] Jeffrey A. Livermore, "Factors that Significantly Impact the Implementation of an Agile Software Development Methodology", Journal of Software, Vol. 3, No. 4, APRIL 2008.

[15] Jennifer Stapleton, "DSDM: The Method in Practice", Addison-Wesley, 1997.

[16] John Hunt, "Agile Software Construction", Springer, 2006.

[17] K.Usha, N.Poonguzhali, and E.Kavitha, "A Quantitative Approach for Evaluating the Effectiveness of Refactoring in Software Development Process", International Conference on Methods and Models in Computer Science, Delhi, India, Dec. 2009.

[18] Karlheinz Kautz and Sabine Zumpe, "Just Enough Structure at the Edge of Chaos: Agile Information System Development in Practice", 9th International Conference, XP 2008, Limerick, Ireland, Proceedings, June 2008.

[19] Kent Beck, "Extreme Programming Explained: Embrace Change", Addison Wesley, 1999.

[20] N. Wallace, P. Bailey, and N. Ashworth, "Managing XP with Multiple or Remote Customers", Third International Conference on eXtreme Programming and Agile Processes in Software Engineering (XP2002), 2002.

[21] Noura Abbas, Andrew M. Gravell, and Gary B. Wills, "Historical Roots of Agile Methods: Where Did Agile Thinking Come From?", 9th International Conference, XP 2008, Limerick, Ireland, Proceedings, June 2008.

[22] Pekka Abrahamsson, Outi Salo, Jussi Ronkainen, and Juhani Warsta, "Agile Software Development Methods – Review and Analysis", VTT, 2002.

[23] R. C. Martin, "Extreme Programming - Development Through Dialog", IEEE Software, pp. 12–13, 2000.

[24] S.R. Palmer and J.M. Felsing, "A Practical Guide to Feature-Driven Development", Prentice-Hall Inc, 2002.

[25] Tobias Hildenbrand, Michael Geisser, Thomas Kude, Denis Bruch, and Thomas Acker, "Agile Methodologies for Distributed Collaborative Development of Enterprise Applications", International Conference on Complex, Intelligent and Software Intensive Systems, 2008.

[26] Williams, L., Kessler, R., Cunningham, W., and Jeffries, R, "Strengthening the Case for Pair Programming", IEEE Software 17, 19–25, 2000.

[27] Yang Yong and Bosheng Zhou, "Evaluating Extreme Programming Effect through System Dynamics Modeling", International Conference on Computational Intelligence and Software Engineering (CiSE), Wuhan, China, Dec. 2009.

Software Complexity Methodologies & Software Security

Masoud Rafighi
Taali University, Iran
Qom, Iran
Masoud_r62@yahoo.com

Nasser Modiri
Faculty Memeber, Zanjan Azad University, Iran
Tehran, Iran
Nassermodiri@yahoo.com

Abstract—**It is broadly clear that complexity is one of the software natural features. Software natural complexity and software requirement functionality are two inseparable part and they have special range. measurement complexity have explained with using the MacCabe and Halsted models and with an example discuss about software complexity in this paper Flow metric information Henry and Kafura, complexity metric system Agresti-card-glass, design metric in item's level have compared and peruse then categorized object oriented and present a model with 4 level of software complexity, we can create a decent understanding of software security best practices that can be practically applied and make a big impact on the software security problem.**

Keywords— McCabe model, Halstead model, measurement software complexity, security software.

I. Introduction

Due to high cost of software, software organization are trying to find away to make it lower. Because of this the researcher are trying to find the relation of software feature and problem of extended software. Hard works need more time to do, in this time we need more sources, that it means more cost. One of the reasons for proceeding to software's complexity and its measurement is controlling the expenditure of software's life time, because software complexity is one of the basic agents in increasing cost of extended and maintenance. Software complexity is an item that is not identified and it's not easy to measure and describe and usually disregarded in planning project process. So we are looking for a way to predict how hard maintenance, change and understanding software is. That with measurement and control decreases the cost on software's life time

Due to high cost of software, software organization are trying to find away to make it lower. Because of this the researcher are trying to find the relation of software feature and problem of extended software. Hard works need more time to do, in this time we need more sources, that it means more cost. One of the reasons for proceeding to software's complexity and its measurement is controlling the expenditure of software's life time, because software complexity is one of the basic agents in increasing cost of extended and maintenance. Software complexity is an item that is not identified and it's not easy to measure and describe and usually disregarded in planning project process. So we are looking for a way to predict how hard maintenance, change and understanding software is. That with measurement and control decreases the cost on software's life time

.

II. Complexity measure

Basic of complexity describe is quality of connection between different part of software system, the simplest metric for structure complexity is measure. The measure determine with LOC or functional point.

✓ LOC

One of the most famous balance software is line counter with LOC unit or about big program with KLOC which is used for quantity of software complexity. Unfortunately there is no agreement on every part of LOC. most of the researcher come to an agreement to not calculate the distance of lines. But yet there is no agreement about comment, sign, and structure like BEGIN in Pascal and...

Another problem in free format language is different structure are in one textual line or one executive structure is broken to more than one line executive code.

LOC metric is simple, understandable; it used in every program language and it has wide usage. Also we can use it for evaluation programmer although it needs attention because of the style of programming it can has effect on values, a programmer it can has effect on values, a programmer may produce many lines and another one be success to compress that function in lower space. Also extender, work on different thing except producing more code, like document, programming test and... also the time of wage payment to code line need more attention because there is many way to make the program massive.

Function point metrics

Quantities metric which are base on the number of code line program are not satisfied. From the user point of view function points are a group of measurable code. A huge program may have millions LOC. But a program with 1000 function points is a huge application program or a real system. A function as a collection of programmable structure, with definition of formal parameter and local variable that change with this structure is defined.

A metric of functionality point, in IBM is a weighted total of five items that characterize a application program.

Function point is coming from a tentative relation base on metric countable from software information domain and evaluation of software complexity.

Function point will caulk with a complete table. Five feature of domain will determine. There are counts in suitable place of table. To determine the values of information domain flow this sentences:

The number of incoming user: every incoming user that has different application data from software will count. Entrance should count different from requests.

The number of outgo user: every outgo user that brings information for user will count. In this paper, outgo is reports, monitor, error massages and...

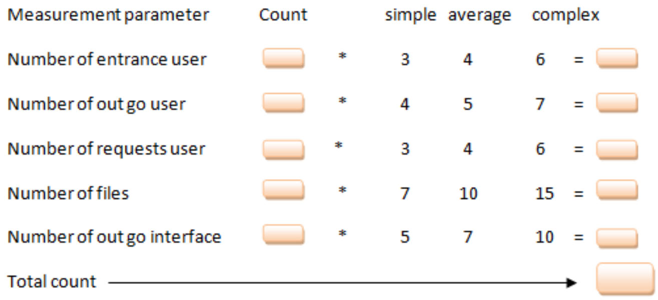
Sporadic ingredient data in a text report, won't count differently.

The number of user's requests: the request will define as a online entrance which produce answer without any pause every one of the requests will count.

The number of files: every main logical files is a logical group of data which can be part of a big information bank or a separate file, and will count.

The number of outgo interface: all of the machine reading (like data file on thin tape) which uses to transfer the information to another system will count.

Weighted coefficient



Measurement parameter	Count		simple	average	complex		
Number of entrance user	□	*	3	4	6	=	□
Number of out go user	□	*	4	5	7	=	□
Number of requests user	□	*	3	4	6	=	□
Number of files	□	*	7	10	15	=	□
Number of out go interface	□	*	5	7	10	=	□
Total count						→	□

Figure 1. Function point.

One complex value will determine for every count when the data has assembled. The organization which use this way will develop determination simple, average or complex portal evidences. For function point (FP) use this frame:

FP = total count x[0.65+0.01x $\sum(F_i)$].

(1)

Total count: sum all FP portals which is in fig.1

F_i (I =1 to 14) <<Value of complexity conduction>> base on answer of these questions:

1. Does system need support and retrieval?
2. Does it need connection data?
3. Is there any parcel processing operation?
4. How important is efficiency?
5. Does system work in a operational environment?
6. Does system need online data portal?
7. Does online data online need to make input transaction on operation or multi job monitor?
8. Does main files update online?
9. Are the entrances, outgoes, files and requests complex?
10. Is the internal process complex?
11. Are the codes usable again?
12. Is there any reduction or installation in design?
13. Is it designed for installing in different organization?
14. Does the application program make the changes simple and use easily by user?

The answer of this question is between 0 to 5, the constant values in this frame have found tentative.

When function points were calculated, they are used in a way like LOC method. For normalization of software implement qualification, quantity and another qualification.

III. Other complexity metrics

✓ Cyclic number McCabe

Cyclic complexity is the most usage member of static software metric. Cyclic complexity measure the number of liner independence way in a yardstick. It shows a number which can compare with other programs complexity. Cyclic complexity is program complexity or McCabe complexity. It's easy to understand this complexity and you can get useful result.

This measure is independent from language and format language. Cyclic number is a simple way to compare software.

Cyclic complexity measure is coming from connection graph to measure.

CC = E - N + p. (2)

E: number of edge graph

N: number of disconnect nod P: number of disconnect part of graph

Countable treaties are needed for real count this item. For example some tools which get cyclic complexity have this treaty. this complex number give you a better measure to calculate the program complexity. this figure show a part of code and connection graph with cyclic number 9.

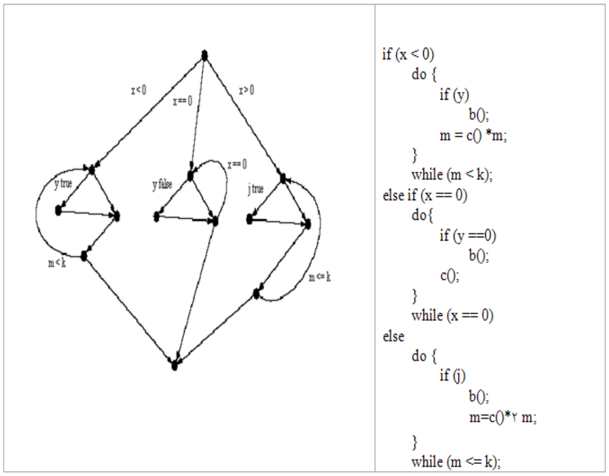
Nodes which have more than one way increase the cyclic complexity.

Another way to calculate cyclomatic complexity is:

Cc= number of decision +1.

(3)

So, what's the decision? Decisions come from conditional predicate. The cyclomatic complexity of a procedure without any decision is 1.there is no maximum value for cyclomatic complexity because one procedure can have many decision. Conditional predicate, include for, case, if ... then.... else..., while, do and...



```
if (x < 0)
      do {
            if (y)
                  b();
            m = c() *m;
      }
      while (m < k);
else if (x == 0)
      do{
            if (y ==0)
                  b();
            c();
      }
      while (x == 0)
else
      do {
            if (j)
                  b();
                  m=c()*۲ m;
      }
      while (m <= k);
```

Figure 2. example of cyclic complexity graph.

Its merit to mention that cyclic complexity is not sensitive about unconditional junction like go to, return and break-statement, however they increase complexity. The complexity of many programs are measure and determine a confine for complexity that help software engineers to find the natural risk and perpetuity of a program.

Table I. Effect of conditional predicate in cyclic complexity

+1	⇨	If…Then
+1	⇨	Else...If..Then
+1	⇨	Case
+1	⇨	For [Each]
+1	⇨	Do
+1	⇨	While

Criterion which is regulated for development and maintenance and for estimate this risk, coast and perpetuity program in reengineering can use. Studies show that the cyclic complexity program and errors frequency are dependent. The low complexity help out to understand program easier. Having changes in programs which are low cyclic complexity have lower risk than programs which are high cyclic complexity. Also cyclic complexity of yardstick is a powerful measure to test it. One common cyclic complexity usage is comparing it with a collection threshold value. You can see this collection in table II.

Table II. Cyclic complexity

CC	Kind of procedure	Risk
1-4	One simple procedure	Low
5-10	One perennial procedure with good structure	Low
11-20	A complex procedure	Average
21-50	A complex warning procedure	High
>50	A susceptible of error and changeable procedure	Very high

cyclic complexity is usage in different precinct like:

- ✓ Analysis code development risk
- ✓ Analysis changes in maintenance risk
- ✓ Test planning
- ✓ Halsted's metric

IV- Halsted metric

Professor Maurice Halstead separates the software knowledge and computer knowledge. Criterion of Halstead complexity for measurement the range of yardstick program complexity is coming from source code. Halstead's criterions were for determine a quantities criterions from yardstick's values. These criterions were the most powerful typical determine the code complexity between primary metrics. This metric use as a maintenance metric to apply the metrics to code. There is much different idea about value of Halstead criterion which is in the range of "complexity... and unreliable" to "the most powerful maintenance criterion". one thing which is so important is reliable to tentative document in typical maintenance, but it's clear that this Halstead criterion are useful even in development state for estimate the quality of code in programs which have high calculative density [1].Halstead's criterions are based on four value which are from code source.

n_2 : Number of different values which are in program.

N_1 : Total number of operator

N_2 : total number of values

This numbers cause 5 criterions:

Table III. Halstead metric

Criterion	Symbol	Frame
Length of program	N	N= N1 + N2
Collection of word program	N	n= n1 + n2
Bulk	V	V= N * (LOG2 n)
Difficulty	D	D= (n1/2) * (N2/n2)
Effort	E	E= D * V

If one time a rule for calculating the value be specified, it's easy to calculate this criterion. Derivation of number of code items needs a sensitive scanner which is a simple program for most of the languages. Halstead's criterions are operational in operational system and for development effort one time after writing the code. Code maintenance at development time have to attend, Halstead's criterions should use during code development the pursuit the complexity. They were criticized duo to difference reasons. This is a claim which says these criterions measure lexical and textual complexity not structural or logical flow complexity. However that the most powerful measure criterions is maintenance. Specially, estimate the complexity with Halstead's criterions for code which has high rate of logic calculations instead of logic junction is tenderer. Cyclic complexity is one of the structural complexity criterions. Another metrics express other aspect of complexity; include structural and calculative complexity as what you see on table IV.

Table IV. Example of criterion of complexity

Criterion of complexity	Usual criterion
Halstead's Criterion of complexity	Algorithmic complexity will measure by counting values
Henry and Kafura metrics	Connection between yardsticks(parameters, public, values, calling)
Bowles metrics	System and yardstick complexity, connecting by parameters and public values
Troy and Zweben metrics	Connection or to be yardstick, structure complexity (maximum depth structure chart) call to, call by
Ligier metrics	To be yardstick structure chart

V. Object-oriented complexity model

Paradigm OO by using a better way to analysis problem, plan and implement solution is basic change in software engineering. Most of the software engineering purposes are accessible like maintenance, reliable, usable.

Some advantages of OO system is fast development, high quality, easy maintenance, decreasing coast, better informational structure and increasing compatibility. One of the main reasons of this claims is OO methods with support of data secession hierarchy analysis.

Some important question which should be answered:

What is the difference between OO paradigm and primary paradigm?

How these differences make access to software engineering purpose easier?

Are this purpose really as they were claimed?

To answer this question we need to have ability measurement and suitable criterion.

Software metrics have many cohort as a basic rule in a engineering way for design and OO software development control like software complexity level.

Complexity of OO system can express with a collection of criterion which define in deferent level. A model of complexity system with four levels has suggested for OO system: values, method, object, system.

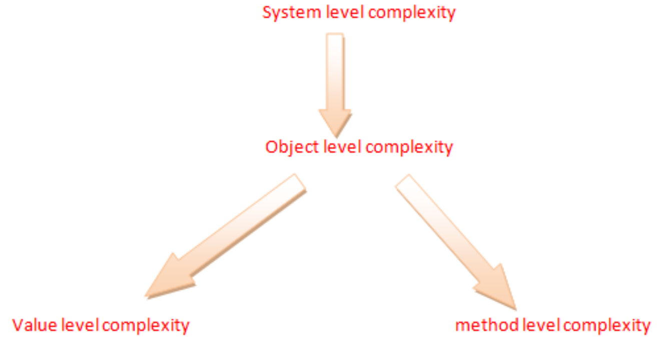


Figure 3. a model of complexity in object-oriented system with 4 level

Value level complexity have relation with definition of values in system method level complexity have relation with definition of method in system object level complexity is a combination of value and method complexity with inheritance structure criterions. System level complexity gives you a performance from high level of organization and size of OO system. There are some criterions to make system connection acceptable in every level. Criterions are usable in every part of systems life OO metrics can be calculated in different levels. We can have some metrics in level of system which assemblage structural feature of all part of system. In class level we can calculate the structural feature of class like union and depth of inheritance. We can determine some metrics on method levels.

VI. Software security

Software security best practices applied to various software artifacts. Although the artifacts are laid out according to a traditional waterfall model in figure 4, most organizations follow an iterative approach today, which means that best practices will be cycled through more than once as the software evolves.

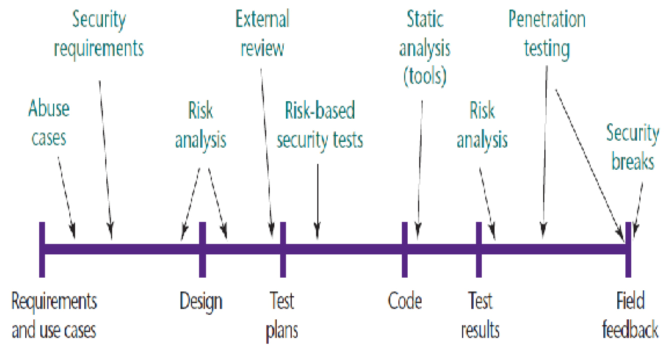


Figure 4 . The artifacts are laid out according to a traditional waterfall model.

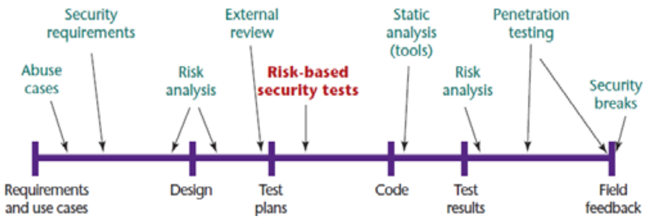


Figure 5 . The software development life cycle.

Throughout this series, we'll focus on specific parts of the cycle; here, we're examining risk-based security testing [7]. There is no silver bullet for software security; even a reasonable security testing regimen is just a start.

Unfortunately security continues to be sold as a product, and most defensive mechanisms on the market do little to address the heart of the problem, which is bad software. Instead, they operate in a reactive mode: don't allow packets to this or that port, watch out for files that include this pattern in them, throw partial packets and oversized packets away without looking at them. Network traffic is not the best way to approach this predicament, because the software that processes the packets is the problem. By using a risk-based approach to software security testing, testing professionals can help solve security problems while software is still in production [8].

6. Conclusions

Software metrics are useful technique. To improve quality we have to find a method to measure the complexity of software for control and supervision on it. In this paper, the algorithms and methods of measurement the software complexity are compared. Studies and researches show that we can find the complexity by using algorithms and different methods as the high level of complexity cause many errors, need to test it and high coast of development and maintenance. so, software complexity has directly relation with coast of development and maintenance. so it's not logical to disregard it. As result to decrease the coast of maintenance and repairing software you should measure and restrain the complexity of software. It is suppose that the present ways to measure the software complexity has wide domain that we should guide it to requirement complexity if we remove complexity sooner. We will have fewer coasts so it's logical to looking for methods to measure the complexity in first phase of software production (requirements phase, analysis and design phase). As the trinity of trouble connectedness, complexity, and extensibility continues to impact software security in a negative way, we must begin to grapple with the problem in a more reasonable fashion. Integrating a decent set of best practices into the software development life cycle is an excellent way to do this. Although software security as a field has much maturing to do, it has much to offer to those practitioners interested in striking at the heart of security problems.

REFERENCES

[1] Sylvia B. Sheppard, Phil Milliman, M. A. Borst, and tom love."Measuring the Psychological Complexity of Software Maintenance Tasks with the Halstead and McCabe Metrics" IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL.

[2] SE-5, NO. 2, MARCH 1979. Pp.96-104
Yas Alsultanny." Using McCabe Method to Compare the Complexity of Object Oriented Languages" IJCSNS International Journal of Computer Science and Network Security,VOL.9 No.3, March 2009.pp.320-326

[3] Paul. D. Scott." Measuring Software Component Reusability by Coupling and Cohesion Metrics" JOURNAL OF COMPUTERS, VOL. 4, NO. 9, SEPTEMBER 2009,797-805

[4] Yingxu Wang and Jingqiu Shao," Measurement of the Cognitive Functional Complexity of Software" Proceedings of the Second IEEE International Conference on Cognitive Informatics (ICCI'03)0-7695-1986-5/03 2003 IEEE

[5] Jitender Kumar Chhabra, K.K. Aggarwal, Yogesh Singh," Code and data spatial complexity: two important software understandability measures" Information and Software Technology 45 (2003) 539–546

[6] S. R. Chidamber and C. F. Kemerer, "A Metrics Suite for Object Oriented Design," IEEE Trans. on Software Eng., vol. 20, no.6, 1994, pp. 476-493.

[7] D. Verndon and G. McGraw, "Risk Analysis in Software Design," *IEEE Security & Privacy*, vol. 2, no. 4, 2004, pp. 79–84.

[8] G. McGraw, "Software Security, "*IEEE Security & Privacy*, vol. 2, no.2, 2004, pp. 80–83.

[9] A. Lapouchnian, S. Liaskos, J. Mylopoulos,
Y. Yu. Towards Requirements-Driven Autonomic Systems Design. In Proc. ICSE 2005 Workshop on Design and Evolution of Autonomic Application Software (DEAS 2005), St. Louis, Missouri, USA, May 21, 2005. ACM SIGSOFT Software Engineering Notes 30(4), July 2005.

AUTHORS PROFILE

Masoud rafighi was born in tehran, Iran on 1983/08/10. he receive M.Sc degree in computer engineering software from Azad University North Tehran Branch, Tehran, IRAN. He has recently been active in software engineering and has developed and taught various software related courses for the Institute and university for Advanced Technology, the University of Iran. His research interests are in software measurement, software complexity, requairement engineering, maintanence software, software security and formal metods of software development. He has written a book on software complexity engineering and published many papers.

Nasser Modiri received the MS degree in MicroElectronics from university of Southampton, UK in 1986. He received PHD degree in Computer Networks from Sussex university of UK in 1989. He is a lecture at department of computer engineering at Islamic Azad University of Zanjan, Iran. His research interests include Network Operation Centres, Framework for Securing Networks, Virtual Organizations, RFID, Product Life Cycle Development and Framework For Securing Networks.

An Improved Energy Aware Hierarchical Routing Protocol In Wireless Sensore Networks

Behzad Homayoufar
Department of Technical and Engineering
Mashhad Branch, Islamic Azad University
Mashhad, Iran
BehzadHomayounfar1@gmail.com

Sayyed majid mazinani
Department of Electrical Engineering
Imam Reza University
Mashhad-Iran
Mazinani@ieee.org

Abstract—**Reducing energy consumption and prolonging network lifetime is an important issue in wireless sensor networks. So this problem has to solve for sensor node energy while meeting the requirements of applications/users. Hierarchical network structures have the advantage of providing scalable and resource efficient solutions. In this paper to find an efficient way for saving energy consumption, we propose an Improved Energy Aware Hierarchical Routing Protocol (IERP) that prolong the sensor network lifetime. IERP introduces a new clustering parameter for cluster head election, routing tree construction on cluster heads for sending aggregated data to the base station. We use two parameters to select cluster heads and construct routing tree on cluster heads that includes distance from each node (others or base station) and residual energy of the nodes. We use a simple but efficient approach, namely, intra-cluster coverage to cope with the area coverage problem. Simulation results in the NS-2 platform demonstrate the longer network lifetime of the IERP than the better-known clustering protocols, ERA and EAP.**

Keywords-Hierachical; Clustring; Routing Tree; Lifetime Network; Residual Energy

I. INTRODUCTION

A typical WSN consists of a number of sensor devices that collaborate with each other to accomplish a common task (e.g. environment monitoring, object tracking, etc.) and report the collected data through wireless interface to a sink node. The areas of applications of WSNs vary from civil, healthcare and environmental to military. Examples of applications include target tracking in battlefields[1], habitat monitoring[2],civil structure monitoring [3], forest fire detection [4] and factory maintenance [5].

Wireless sensor networks (WSNs) become an invaluable research area by providing a connection between the world of nature and that of computation by digitizing certain useful information. In wireless sensor networks, the sensor node resources are limited in terms of processing capability, wireless bandwidth, battery power and storage space, which distinguishes wireless sensor networks from traditional ad hoc networks [6]. In most applications, each sensor node is usually powered by a battery and expected to work for several months to one year without recharging. Such an expectation cannot be achieved without carefully scheduling the energy utilization. So one of the very important factors that effect on sensor network life time is sensor's energies, so the protocol running on sensor networks must efficiently reduce the energy consumption in order to prolong network lifetime [7]. Data gathering is a typical operation in many WSN applications, and data aggregation in a hierarchical manner is widely used for prolonging network lifetime. Data aggregation can eliminate data redundancy and reduce the communication load. Hierarchical mechanisms (especially clustering algorithms) are helpful to reduce data latency and increase network scalability [8]. IERP protocol introduce new formula for cluster head selection that can better handle homogeneous energy circumstances than other clustering algorithms which IERP, first cluster the network then construct a spanning routing tree over all of the cluster heads. IERP uses two parameters to select heads on tree that includes distance from each node (others and base station) and residual energy of the nodes. Only the root node of this tree can communicate with the sink node by single -hop communication. Because the energy consumed for all communications in network can be computed by the free space model, the energy will be extremely saved and Network lifetime is prolonged. The rest of this paper is organized as follows: In the next section we introduce the related work, in section 3 we will discuss the proposed algorithm, simulation results and performance evaluation are given in section 4, the conclusion is presented in sections 5.

II. RELATED WORKS

In hierarchical networks, nodes are separated to play different roles, such as CHs and cluster members. The higher level nodes, cluster heads (CHs), Each CH collects data from the cluster members within its cluster, aggregates the data, and then transmits the aggregated data to the sink. All of the hierarchical routing protocols aim at selecting the best CH and clustering the nodes into appropriate clusters in order to save energy. The hierarchical clustering protocol may execute reclustering and reselecting of CHs periodically in order to distribute the load uniformly among the whole network [10]. By the method of CH selection, the hierarchical routing protocols can be classified into two categories: random-selected-CH protocol and well-selected- CH protocol. The former randomly selects CHs and then rotates the CH task among all nodes, while the latter carefully selects appropriate

CHs and then gathers nodes under the CHs based on the network status [9] and [10]. Energy Residue Aware (ERA) clustering algorithm is one of energy-aware hierarchical approaches. It is also improved from LEACH by including the communication cost into the clustering. The communication cost includes residual energy, communication energy from the CH to the sink and communication energy from the cluster members to the CH. ERA uses the same CH selection scheme as LEACH but provides an improved scheme to help non-CH nodes choose a better CH to join by calculating the clustering cost and finding CH according to maximum residual energy [11].

In HEED, author introduces a variable known as cluster radius which defines the transmission power to be used for intra-cluster broadcast [12]. The initial probability for each node to become a tentative cluster head depends on its residual energy, and final heads are selected according to the intra-cluster communication cost. HEED terminates within a constant number of iterations, and achieves fairly uniform distribution of cluster heads across the network. In EAP(Energy-Aware Routing Protocol), a node with a high ratio of residual energy to the average residual energy of all the neighbour nodes in its cluster range will have a large probability to become the cluster head. This can better handle heterogeneous energy circumstances than existing clustering algorithms which elect the cluster head only based on a node’s own residual energy. After the cluster formation phase, EAP constructs a spanning tree over the set of cluster heads [13]. Only the root node of this tree can communicate with the sink node by single-hop communication. Because the energy consumed for all communications in the network can be computed by the free space model, the energy will be extremely saved and thus leading to sensor network longevity [14].

III. The Proposed Algorithm

In IERP , the role of the cluster head must be rotated among all sensor nodes. Therefore, the operation of IERP is divided into rounds. Each round begins with a set-up phase while clusters are organized and then in the steady-state phase the routing tree is constructed as well as aggregated data are sent to the sink node.

In IERP protocol, each node needs to maintain a neighbourhood table to store the information about its neighbours that including residual energy and distance to sink.

A. *Network Model*

This paper assumes that *N* sensor nodes are randomly scattered in a two-dimensional square field *A* and the sensor network has the following properties:

- This network is a static densely deployed network. It means a large number of sensor nodes are densely deployed in a two-dimensional geographic space, forming a network and these nodes do not move any more after deployment.
- There is only one base station, which is deployed at a fixed place outside A.
- The energy of sensor nodes cannot be recharged.
- Sensor nodes are location-aware, i.e. a sensor node can get its location information through other mechanism such as GPS or position algorithms.

B. *Set-up phase*

At the beginning of each round, each node first estimates its residual energy *$(E_{node\text{-}res})_j$* and broadcasts the *CH-E_Msg* within radio range r which contains residual energy and distance to base station. Each node receives the *CH-E _Msg* from all neighbours in its cluster range and updates the neighbourhood table, also compute *CH-E* (cluster head election) using (1).

$$CH - E = \frac{(E_{node-res})_j}{(1 - (\frac{dis(j)}{100}))^2} \qquad (1)$$

$(E_{NODE\text{-}RES})_J$ can be derived as below:

$$(E_{node-res})_j = Max\{(E_{node-rem})_j - (E_{toOther})_{ji}\} \qquad (2)$$
$$j \in N, \forall i \in S_o$$

Where, N is the set of nodes , S_O is set of other nodes within radio range r and *$(E_{node\text{-}rem})_j$* indicates the residual energy of node j in the current round as well as *$(E_{toOther})_{ji}$* indicates the communication energy from node j to other nodes i within radio range *r*. Eventually, each node chooses *$(E_{node\text{-}res})$* according to maximum residual energy .

Value of parameter *dis(j)* is computed as follow :

$$dis(j) = (\sum_{i=1}^{l} (| D_{db}(j) - D_{db}(i) |) \times t_p \times k) \qquad (3)$$

D_{db} is node distance to base station. We assume that number of bits , *k=1* , Transmission power , *t_p =1*.

In this protocol , If node s *CH-E* is the largest value within radio range *r* , it will set its state as head and node which has the second largest value of *CH-E* is selected as the back up cluster head for the next round. Because , the probability that this node will be selected as cluster head in the next round is high. So minimizing communication energy , calculations of CHs for half of rounds and reduction of energy Consumption for each round can help to prolong the network lifetime.

C. *Construction of Routing Tree*

There are several ways that can construct aggregation tree[16]. All tree algorithms have the same structure but have different metrics and cost measures. In this paper we use two parameters to select root node on tree which is distance from each node (others or base station) and residual energy of the nodes. Only the root node of this tree can communicate with the sink node by single -hop communication. In IERP , After clustering, cluster heads broadcast within a radio range *R* a message contains node residual energy and its distance to base station. The cluster head computes *RN* (root node) by Using (4):

$$RN=\frac{(E_{CH-res})_j}{\sum_{i=1}^{l}\left(\frac{(E_{CH-res})_i}{D_{CH-db}(i)}\right)+\sum_{i=1}^{l}dis(ij)^2}\qquad(4)$$

Where, *(E_{CH-res})* is obtained as follow:

$$(E_{CH-res})_j=(E_{CH-rem})_j-(E_{CH-BS})_j\qquad(5)$$
$$j\ \in S_C$$

S_C is set of cluster heads in radio range R , *(E_{CH-res})$_j$* indicates the residual energy of the cluster head , *(D_{CH-db})* indicates cluster head distance to base station and *dis(ij)* determines distance between cluster heads in radio range *R*.

Each cluster head node compute this *RN* and broadcasts it to other cluster head nodes within its radio range R . If the other cluster head node has smaller *RN* , it selects the node that has the largest *RN* as its parents and sends a message to notify the parent node. Finally, after a specified time, a routing tree will be constructed, whose root node has the largest *RN* among all cluster heads. Example of network topology is shown in Fig. 1.

TABLE I. SIMULATION PARAMETERS

Parameters	Value
Network Filed	(0,0)~(100,100)
Number of nodes	100~500
Cluster radius R	30 m
Sensing radius r	10 m
Sink position	(50,200)
Initial energy	3 J
Data packet size	600 Bytes
Broadcast packet size	30 Bytes
Ethreshold	0.01 J
Eelec	50 nJ/bit
Efs	10 nJ/bit/m2
Threshold distance	80m
Data Cycles per round(L)	5

D. *Intra-Cluster Coverage*

Coverage is one of the most important issues in WSNs and it has been studied extensively in recent years [17]. Coverage mechanism is to choose a subset of active nodes to maintain the coverage expectation. We introduce into clusters the notion of intra-cluster coverage which selects some active nodes within clusters while maintaining coverage expectation of the cluster. Utilizing the idea proposed in our research, cluster head randomly chooses m' nodes according to (6) :

$$p_{cov\,er}=\sum_{i=k}^{m'}C_{m'}^{i}\left(\frac{r}{R}\right)^{2i}\left(1-\frac{r^2}{R^2}\right)^{m'-i}\qquad(6)$$

Where, P_{cover} is the coverage expectation of sensing field, and *r* is sensing radius, *R* is cluster radius and *m'* is the number of active nodes. Use of intra-cluster coverage has two advantages. The first is to reduce energy consumption in each round by turning redundant nodes' radio off so that network lifetime is prolonged. The second is to reduce TDMA schedule overhead. In this case we can coverage whole of network by active nodes and other member nodes are turned off, as a result, energy consumption in intra cluster nodes remarkably reduced and network lifetime is extended [15].

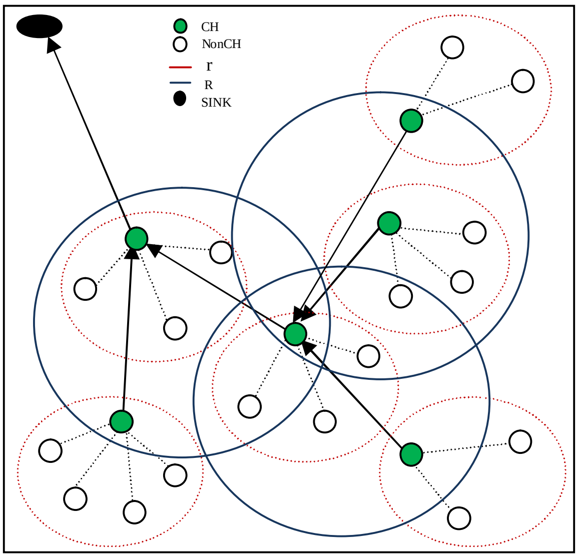


Figure 1. Example of Network Topology

IV. PERFORMANCE EVALUATION

We used NS-2 to implement and simulate our protocol and compare it with the ERA and EAP protocols. Every simulation result shown below is the average of 100 independent experiments where each experiment uses a different randomly-generated uniform topology of sensor nodes. The parameters used in simulations are listed in Table 1.

A. *Network Lifetime*

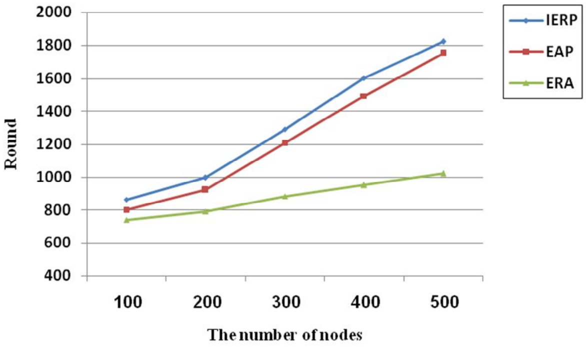


Figure 2. Network Lifetime

Fig.2 shows the network lifetime between ERA, EAP, and IERP protocols with the number of nodes from 100 to 500. As seen in figure, number of rounds is significantly extended due to the reasons .First Cluster head roles are rotated, so energy consumption among cluster members is balanced. Second, constructing routing tree on cluster heads to send aggregated data to the base station as multi-hop that can extremely reduces energy consumption in Cluster heads.

B. Network Lifetime Versus Base station position

As you know in ERA cluster heads, directly communicate with the sink node, the energy consumption for each cluster head is different because the distance between each node and the sink node is different. As a result, energy consumption farthest CHs to the BS more than nearest CHs. So, their energy significantly reduced and nodes die soon. In IERP and EAP protocols, there is only a single node to communicate with the sink node, Fig.3 shows, the network lifetime of three protocols, by changing base station position.

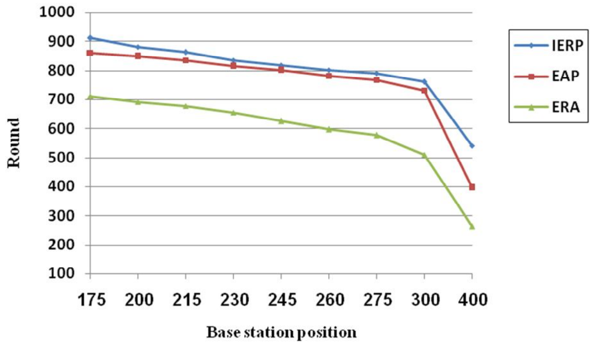


Figure 3. Network Lifetime vs. BS Position

C. Average Energy Consumption in Cluster Heads

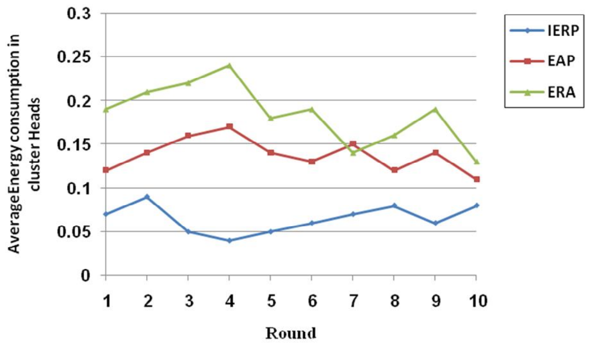


Figure 4. Averag energy consumption in CHs

Energy consumption by cluster heads per round in IERP is lower than that in ERA and EAP Because in ERA cluster heads send their data directly to the Base Station. EAP don't use distance (to base station or other node) for its cluster head election while IERP construct cluster heads and spanning tree on cluster heads based on distance and residual energy and cluster heads send their data in multi hop to the base station so, extremely reduce energy consumption in CHs, as shown in Fig.4 for 10 rounds.

D. Time of the Nodes dead

Fig.5 shows an influence of network topology. we change the number of nodes from 100 to 500 and observe the time of the nodes dead. In ERA and EAP each node has to spend more energy to communicate with other nodes and manage the cluster so the network lifetime decreases with the scale of network while IERP is improved on average time of 100% nodes dead when the number of nodes is changed from 100 to 500.Because, each node has the lower energy consumption.

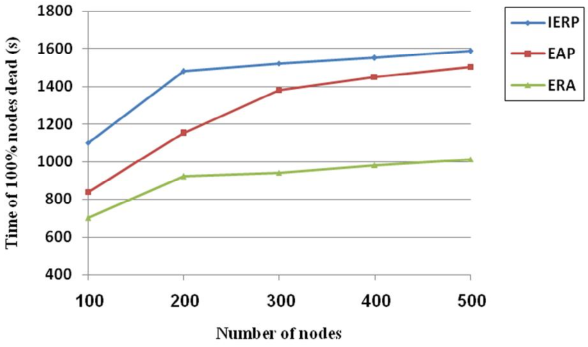


Figure 5. Time of the 100% nodes dead

V. Conclusion

In this paper, to maximize the network lifetime we used hierarchical mechanism with new factors for selecting cluster heads and root node on the tree. Also we introduced new coverage schema for energy saving in member sensors, which can save extremely energy in sensors. According Simulation results, IERP has improved the network lifetime by reducing energy consumption on cluster heads and other sensor nodes, when compared to other protocols.

References

[1] T. Bokareva, W. Hu, S. Kanhere, B. Ristic, N. Gordon, T. Bessell, M. Rutten, S. Jha, "Wireless sensor networks for battlefield surveillance", in: Proceedings of The Land Warfare Conference, LWC Brisbane, Australia, October 24_27, 2006.

[2] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, J. Anderson, "Wireless sensor networks for habitat monitoring", in: The Proceedings of the 1st ACMInternational Workshop on Wireless Sensor Networks and Applications, ACMWSNA, Atlanta, Georgia, USA, September 28_28, 2002, pp. 88_97.

[3] N. Xu, S. Rangwala, K. Chintalapudi, D. Ganesan, A. Broad, R. Govindan, D. Estrin, "A wireless sensor network for structural monitoring", in: The Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems, Baltimore, MD, USA, November 03_05, 2004, pp. 13_24.

[4] M. Hefeeda, M. Bagheri, "Wireless sensor networks for early detection of forest fires", in: The Proceedings of IEEE International Conference on Mobile Adhoc and Sensor Systems, MASS-2007, Pisa, Italy, October 8_11, 2007, pp. 1_6.

[5] K. Srinivasan, M. Ndoh, H. Nie, H. Xia, K. Kaluri, D. Ingraham, "Wireless technologies for condition-based maintenance (CBM) in petroleum plants" in: The Proceeding of the International Conference on

Distributed Computing in Sensor Systems, DCOSS'05, (Poster Session), Marina del Rey, CA, USA, June 30_July 1, 2005, pp. 389_390

[6] Akyildiz, I.F.; Su, W.; Sankarasubramaniam, Y.; Cayirci, E. “Wireless Sensor Network: a Survey”. Comput. Netw. 2002, 38, 392-422.

[7] F .Akyildiz et al., “Wireless sensor networks :a survey”, Computer Networks”, Vol .38, pp .393-422, march 2002.

[8] Pottie, G.J.; Kaiser, W.J. “Wireless Integrated Network Sensors”. Commun. ACM 2000, 43, 51–58.

[9] J.N. Al-Karaki, A.E. Kamal, “Routing techniques in wireless sensor networks: a survey”, IEEE Wireless Commun. 11 (6) (2004) 628

[10] Chung-Horng Lung , Chenjuan Zhou , “Using hierarchical agglomerative clustering in wireless sensor networks:An energy-efficient and flexible approach”, ,in:Department of Systems and Computer Engineering Carleton University, Ottawa, Ontario, Canada K1S 5B6-2010

[11] H. Chen, C.S. Wu, Y.S. Chu, C.C. Cheng, L.K. Tsai, “Energy residue aware (ERA) clustering algorithm for leach-based wireless sensor networks”, in: 2nd International Conference ICSNC, Cap Esterel, French Riviera, France, August 2007, p. 40.

[12] Younis, O.; Fahmy, S. “HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks”. IEEE Trans. Mob. Comput. 2004, 3, 366-379

[13] Ming Liu , Jiannong Cao , Guihai Chen and Xiaomin Wang , “An Energy-Aware Routing Protocol in Wireless Sensor Networks” , in: ISSN 1424-8220 , Sensors 2009,

[14] W .R .Heinzelman, A .Chandrakasan, and H .Balakrishnan, “Energy-Efficient Communication Protocol for Wireless Microsensor Networks”,Proc .of the Hawaii International Conference on System Science, Jan .2000.

[15] Liu, M.; Cao, J. A “Distributed Energy-Efficient data Gathering and aggregation Protocol forWireless sensor networks”. J. Software 2005, 16, 2106-2116.

[16] Kilhung Lee: “An Energy-Aware Aggregation Tree Scheme in Sensor Networks” in: IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.5, May 2008

[17] Gao, Y.; Wu, K.; Li, F. “Analysis on the redundancy of wireless sensor networks”. In Proceedings of the 2nd ACM international conference on Wireless sensor networks and applications (WSNA 03), September 2003, San Diego, CA, 2003; 108-114.

Java-Based Intrusion Detection System in a Wired Network

Eugène C. Ezin #1, Hervé Akakpo Djihountry #2

# Institut de Mathematiques et de Sciences Physiques
Unité de Recherche en Informatique et Sciences Appliquees
University of Abomey-Calavi
BP 613 Porto-Novo, Republic of Benin
1 eugene.ezin@imsp-uac.org
2 herve.akakpo@imsp-uac.org

***Abstract*—Intrusion Detection has become an integral part of the information security process. The cost involved in protecting network resources is often neglected when compared with the actual cost of a successful intrusion, which strengthens the need to develop more powerful intrusion detection systems. Many existing systems for intrusion detection are developed in C, Objective-C, Tcl, C++ programming languages.**

In this paper, we design and develop a network intrusion detection system using Java programming language. We simulate the land attack, the flooding attack and the death's ping attack to show the effectiveness of the proposed system in which packets in the network are captured online as they come on the network interface.

***Keywords-component*—Intrusion Detection System (IDS), JpCap library, Network Security.**

I. Introduction

With the proliferation of networked computers and the Internet, their security has become a primary concern. This rapid advancement in the network technologies includes higher bandwidths and ease of connectivity of wireless and mobile devices. In 1980, Anderson proposed that audit trails should be used to monitor threats [1]. The importance of such data was not been understood at that time and all the available system security procedures were focused on denying access to sensitive data from an unauthorized source. Latter, Dorothy [2] proposed the concept of intrusion detection as a solution to the problem of providing a sense of security in computer systems. This intrusion detection model is independent of system, type of intrusion and application environment.

Intrusion detection according to Bace is the process of intelligently monitoring the events occuring in a computer system or network, analyzing them for signs of violations of the security policy [3]. In short, intrusion detection is the process of monitoring computers or networks for unauthorized entrance, activity, or file modification. Intrusion detection systems refer to those systems which are designed to monitor an agent's activity to determine if the agent is exhibiting unexpected behavior. Intrusion detection model was proposed by Denning [2]. A more precise definition is found in [4] in which an intrusion detection system is a system that attempts to identify *intrusions*, which we define to be unauthorized uses, misuses, or abuses of computer systems by either authorized users or external perpetrators. Some intrusion detection systems monitor a single computer, while others monitor several computers connected by a network.

Intrusion detection systems detect intrusions by analyzing information about user activities from sources such as audit records, system tables, and network traffic summaries. In short, intrusion detection systems can also be used to monitor network traffic, thereby detecting if a system is being targeted by a network attack such as a denial of service attack.

The primary aim of intrusion detection system is to protect the availability, confidentiality and integrity of crytical networked information systems. Intrusion detection systems are defined by both the method used to detect attacks and the placement of the intrusion detection system on the network. The objective of an intrusion detection system is to provide data security and ensure continuity of services provided by a network [5].

Two major approaches are used by intrusion detection systems: misuse detection and anomaly detection.

Intrusion detection system may perform either misuse detection or anomaly detection and may be deployed as either a network-based system or a host-based system. This description of intrusion detection system leads to four general groups: misuse-host, misuse-network, anomaly-host, and anomaly-network.

Some intrusion detection systems combine qualities from all these categories by implementing both misuse and anomaly detection, and are known in literature as hybrid systems [6]. Even though Gupta in [7] gives an overview on robust and efficient intrusion detection systems, the intrusion detection problem is a hard one since no security is absolutely guarantee for ever.

The goal of this paper is to propose a model for intrusion detection with three different positions for the intrusion detection system using Java programming language. The *Jpcap* library is used in the implementation. So doing, the overall system has more chance to detect an attack. To show the effectiveness of the overall system, three different attacks are simulated.

The paper is organized as follows: section II presents different phases of an attack. Section III gives an overview on the two approaches to intrusion detection. Section IV presents some intrusion detection systems. Section V presents the design of the intrusion detection system we proposed through subsection V-A which describes the functional components of the authentification process. Subsection V-B describes the functional description of the proposed system. Architectures and possible locations of the proposed network intrusion detection system are given in subsection V-D. A description of the plateform is given in section V-E while section V-F describes the involved open source tools to realize the network intrusion detection system. Section VI presents the global architecture.

II. Types of Attack

Classes of attack might include passive monitoring of communications, active network attacks, close-in attacks, exploitation by insiders, and attacks through the service provider. Information systems and networks offer attractive targets and should be resistant to attack from the full range of threat agents, from hackers to nation-states. A system must be able to limit damage and recover rapidly when attacks occur. There are eleven types of attack namely: passive attack, active attack, distributed attack, insider attack, close-in attack, phishing attack, password attack, buffer overflow attack, hijack attack, spoofing attack, exploit attack.

A. *Passive Attack*

A passive attack monitors unencrypted traffic and looks for clear-text passwords and sensitive information that can be used in other types of attacks. Passive attacks include traffic analysis, monitoring of unprotected communications, decrypting weakly-encrypted traffic, and turing authentification information such as passwords. Passive interception of network operations enables adversaries to see upcoming actions. Passive attacks result in the disclosure of information or data files to an attacker without the consent or knowledge of the user.

B. *Active Attack*

In an active attack, the attacker tries to bypass or break into secured systems. This can be done through stealth, viruses, worms, or Trojan horses. Active attacks include attempts to circumvent or break protection features, to introduce malicious code, and to steal or modify information. These attacks are mounted against a network backbone, exploit information in transit, electronically penetrate an enclave, or attack an authorized remote user during an attempt to connect to an enclave. Active attacks result in the disclosure or dissemination of data files, deny of service, or modification of data.

C. *Distributed Attack*

A distributed attack requires that the adversary introduce code, such as a Trojan horse or back-door program, to a *trusted* component or software that will later be distributed to many other companies and users. Distribution attacks focus on the malicious modification of hardware or software at the factory or during distribution. These attacks introduce malicious code such as a back door to a product to gain unauthorized access to information or to a system function at a later date.

D. *Inside Attack*

An insider attack involves someone from inside, such as a disgruntled employee, attacking the network. Insider attacks can be malicious or not. Malicious insiders intentionally eavesdrop, steal, or damage information; use information in a fraudulent manner; or deny access to other authorized users. No malicious attacks typically result from carelessness, lack of knowledge, or intentional circumvention of security for such reasons as performing a task.

E. *Close-In Attack*

A close-in attack involves someone attempting to get physically close to network components, data, and systems in order to learn more about a network. Close-in attacks consist of regular individuals attaining close physical proximity to networks, systems, or facilities for the purpose of modifying, gathering, or denying access to information. Close physical proximity is achieved through surreptitious entry into the network, open access, or both.

One popular form of close-in attack is social engineering in a social engineering attack, the attacker compromises the network or system through social interaction with a person, through an electronic mail or phone. Various tricks can be used by the individual to reveal information about the security of company. The information that the victim reveals to the hacker would most likely be used in a subsequent attack to gain unauthorized access to a system or network.

F. *Phishing Attack*

In phishing attack the hacker creates a fake web site that looks exactly like a popular site. The phishing part of the attack is that the hacker then sends an e-mail message trying to trick the user into clicking a link that leads to the fake site. When the user attempts to log on with their account information, the hacker records the username and password and then tries that information on the real site.

G. *Password Attack*

In a password attack an attacker tries to crack the passwords stored in a network account database or a password-protected file. There are three major types of password attacks: a dictionary attack, a brute-force attack, and a hybrid attack. A dictionary attack uses a word list file, which is a list of potential passwords. A brute-force attack is when the attacker tries every possible combination of characters.

H. *Buffer Overflow Attack*

Buffer overflow attack is produced when the attacker sends more data to an application than is expected. A buffer overflow attack usually results in the attacker gaining administrative access to the system in a command prompt or shell.

I. Hijack Attack

In a hijack attack, a hacker takes over a session between you and another individual and disconnects the other individual from the communication. You still believe that you are talking to the original party and may send private information to the hacker by accident.

J. Spoofing Attack

In a spoofing attack, the hacker modifies the source address of the packets he or she is sending so that they appear to be coming from someone else. This may be an attempt to bypass firewall rules.

K. Exploit Attack

In this type of attack, the attacker knows a security problem within an operating system or a piece of software and leverages that knowledge by exploiting the vulnerability.

III. Different Approaches to Intrusion Detection

Many classifications exist in literature about intrusion detection [7], [8].

The basic types of intrusion detection are host-based and network-based. Host-based systems were the first type of intrusion detection systems to be developed and implemented. These systems collect and analyze data that originate in a computer that hosts a service, such as a Web server. Once this data is aggregated for a given computer, it can either be analyzed locally or sent to a separate/central analysis machine. Instead of monitoring the activities that take place on a particular network, network-based intrusion detection analyzes data packets that travel over the actual network. These packets are examined and sometimes compared with empirical data to verify their nature: malicious or benign. Because they are responsible for monitoring a network, rather than a single host, network-based intrusion detection systems tend to be more distributed than host-based intrusion detection system. The two types of intrusion detection systems differ significantly from each other, but complement one another well. The network architecture of host-based is agent-based, which means that a software agent resides on each of the hosts that will be governed by the system. In addition, more efficient host-based intrusion detection systems are capable of monitoring and collecting system audit trails in real time as well as on a scheduled basis, thus distributing both CPU utilization and network overhead and providing for a flexible means of security administration.

Two other approaches encountered in literature concerning intrusion detection systems for detecting intrusive behavior are misuse detection and anomaly detection.

A. Misuse Detection

Misuse detection relies on matching known patterns of hostile activity against databases of past attacks. They are highly effective at identifying known attacks and vulnerabilities, but rather poor at identifyning new security threats. Misuse-detection based intrusion detection systems can only detect known attacks.

In [9], the following advantages and disadvantages of misuse detectors can be found.

1) Advantages of misuse detectors: misuse detectors are very efficient at detecting attacks without signaling false alarms. They can quickly detect specially-designed intrusion tools and techniques and provide systems' administrators an easy tool to monitor their systems even if they are not security experts.

2) Disadvantages of misuse detectors: misuse detectors can only detect attacks known beforehand. For this reason the systems must be updated with newly discovered attack signatures. Misuse detectors are designed to detect attacks that have signatures introduced to the system only. When a well-known attack is changed slightly and a variant of that attack is obtained, the detector is unable to detect this variant of the same attack.

B. Anomaly Detection

Anomaly detection will search for something rare or unsual by applying statistical measures or artificial intelligence to compare current activity against historical knowledge. Common problems with anomaly-based systems are that, they often require extensive training data for artificial learning algorithms, and they tend to be more computaionnaly expensive, because several metrics are often maintained, and these need to be updated against every system's activites. Several approaches apply artificial neural networks in the intrusion detection system that has been proposed [10].

Anomaly detection based intrusion detection systems can detect known attacks and new attacks by using heuristic methods.

Anomaly detection-based intrusion detection systems are separated into many sub-categories in the literature including statistical methodologies [11] data mining [12], artificial neural networks [13], genetic algorithms [14] and immune systems [15]. Among these sub-categories, statistical methods are the most commonly used ones in order to detect intrusions by analyzing abnormal activities occurring in the network.

In [9], advantages and disadvantages of misuse detectors can be found.

1) Advantages of anomaly detection: anomaly-based intrusion detection systems, superior to signature-based ones, are able to detect attacks even when detailed information of the attack does not exist. Anomaly-based detectors can be used to obtain signature information used by misuse-based intrusion detection systems.

2) Disadvantages of anomaly detection: anomaly-based intrusion detection systems generally flag many false alarms just because user and network behavior are not always known beforehand. Anomaly-based approach requires a large set of training data that consist of system event log in order to construct a normal behavior profile.

C. Hybrid Intrusion Detection

The hybrid intrusion detection system is obtained by combining packet header anomaly detection and network traffic anomaly detection which are anomaly-based intrusion detection systems with the misuse-based intrusion detection system. Snort is an example of an open-source project for hybrid intrusion detection. The hybrid intrusion detection system is said to be more powerful than the signature-based on its own because it uses the advantages of anomaly-based approach for detecting unknown attacks [9].

IV. Presentation of some Intrusion Detection Systems

There are many implemented intrusion detection systems around the world. Sobirey web site [16] presents more than ninety intrusion detection systems. Some are proprietary (free or commercial) and others are open source. Commercial intrusion detection systems belong to specialized societies in network security such as Cisco System, Computer Associates, Intrusion.com, Network Associates, etc. In the following subsections, we will present some open source intrusion detection systems such as HIDS OSSEC, HIDS Samhain, NIDS Snort, NIDS BRO, IDS Prelude. This choice is motivated by the fact that intrusion detection system we developed is open source using Java technologies.

A. HIDS OSSEC

OSSEC which stands for *open source security* is an open source host-based intrusion detection system. It performs log analysis, file integrity checking, policy monitoring, rootkit detection, real-time alerting and active response. It was initially developed to analyze journal files on servers. Nowadays, OSSEC is able to analyze different journal file formats such as those of Apache, syslog, snort.

B. HIDS Samhain

The Samhain host-based intrusion detection system (HIDS) provides file integrity checking and log file monitoring/analysis, as well as rootkit detection, port monitoring, detection of rogue SUID executables, and hidden processes. Samhain been designed to monitor multiple hosts with potentially different operating systems, providing centralized logging and maintenance, although it can also be used as a stand-alone application on a single host. Samhain is an open-source multiplatform application for POSIX systems (Unix, Linux, Cygwin/Windows).

C. NIDS Snort

Snort is the most commonly used signature-based intrusion detection system and the most downloaded. It is a fast, signature-based and open-source intrusion detection system which produces alarms using misuse rules. It uses binary tcpdump-formatted files or plain text files to capture network packets. Tcpdump is a software program that captures network packets from computer networks and stores them in tcpdump-formatted files. Snort has a language to define new rules. Snort is an open-source project and it has an architecture making it possible to integrate new functionalities at the time of compilation [17], [18].

D. NIDS BRO

Bro is an open source Unix based network intrusion detection system [19]. It is a stand-alone system for detecting network intruders in real-time by passively monitoring a network link over which the intruder's traffic transits. Bro is conceptually divided into an *event engine* that reduces a stream of (filtered) packets to a stream of higher-level network events, and an interpreter for a specialized language that is used to express a site's security policy.

E. IDS Prelude

Prelude has a modular architecture and is distributed. Modular, because its components are independent, and can be easily updated. Distributed, because these independent components interact with each other. This allows to have different components installed on various machines and to reduce the overloaded applications. These various components are the probes and the managers. The probes can be of two types: network or room. A probe network analyzes all the traffic, to detect possible signatures' attacks. The local probe ensures the monitoring of only one machine, and it analyzes the system's behavior to detect attempts of internal vulnerabilities. The probes announce the attempts of attacks by alarms. These alarms are received by the manager who interprets and stores them.

V. Description of the Proposed Design of Intrusion Detection System

This description concerns the authentification process and the network intrusion detection system proposed.

A. Functional Description of the Authentification Process

The system administrator requests for connection to the proposed network intrusion detection system. After three unsuccessful tests the system is disconnected. The following sequences must be carried out:

- the system presents the authentification form,
- the administrator enters his/her login and password,
- the system checks the login and the password,
- the system allows the administrator to have an access to the proposed network intrusion detection or the system doesn't allow the administrator after three unfruitful tests.

|Figure 1 presents the identification process of the system administrator.

B. Functional Description of the NIDS Proposed

When the authentification occurs successfully, the graphical interface of the network intrusion detection system proposed is posted. The following sequences must be then carried out:

- request for choice of an interface network by the administrator,
- posting of the interfaces available on the system;

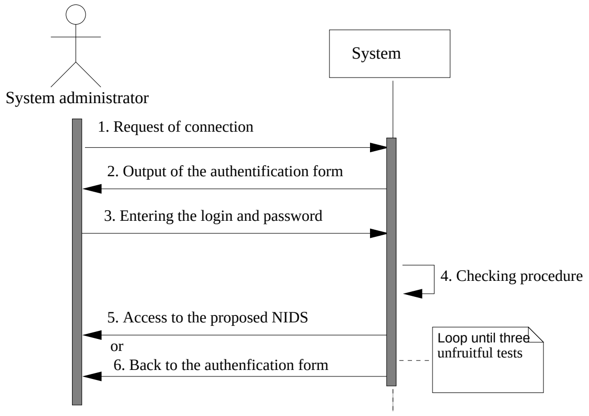

System administrator
System
1. Request of connection
2. Output of the authentification form
3. Entering the login and password
4. Checking procedure
5. Access to the proposed NIDS
or
6. Back to the authenfication form
Loop until three unfruitful tests

Fig. 1. Functional description of the proposed network intrusion detection.

- choice of the interface followed by the network packets capturing process,
- capturing network packets and analyzing specifically of the aforesaid packets,
- alarm's generation as soon as an intrusion is detected,
- querying the database,
- heuristic analysis,
- generating the alarms.
- recording alarms,
- recording of the packets.

Figure 2 presents details about the functional description on the proposed network intrusion detection system.

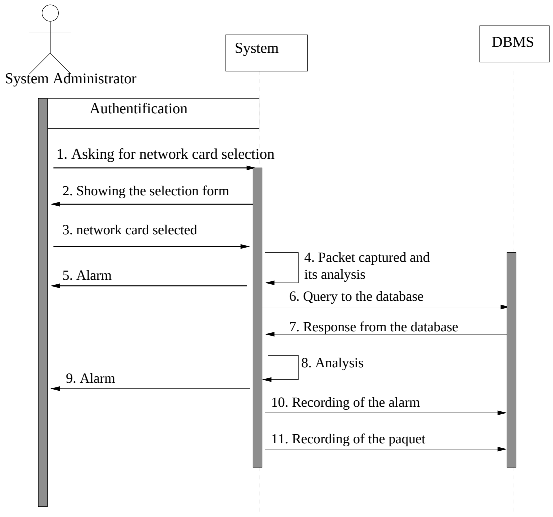

System Administrator
System
DBMS
Authentification
1. Asking for network card selection
2. Showing the selection form
3. network card selected
4. Packet captured and its analysis
5. Alarm
6. Query to the database
7. Response from the database
8. Analysis
9. Alarm
10. Recording of the alarm
11. Recording of the paquet

Fig. 2. Functional description of the proposed network intrusion detection.

C. Attacks in Concern by the implemented System

The proposed network intrusion detection system is intended to detect numerous attacks. Since it is not possible to design an intrusion detection system for every type of attack, we design it for deny of service attack, Web server attack, buffer overflow attack.

D. Architecture and location of the Network Intrusion Detection Systems

The proposed architecture of the network intrusion detection is depicted in Figure 3.

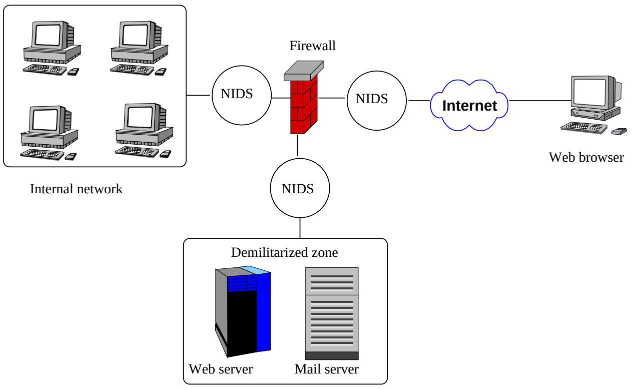


Fig. 3. Proposed architecture and different locations of the proposed network intrusion detection system.

E. Plateform Description

The network intrusion detection we developed is tested on x86 architecture machines. It is also possible to run it in other plateforms. The programming language chosen is Java. This is motivated by little literature in the field of network instrusion detection development in such a language. Many existing intrusion detection systems are developed in C, Objective-C, C++, Tcl.

F. Presentation of the Open Source Tools Used

Many open source tools are used to implement the network intrusion detection system we are proposing. Among them WinPcap, JpCap, JavaMail, MySQL. The following subsections give an overiew on each of them.

1) Presentation of the WinPcap: Packet CAPture is a programming interface that allows to capture the traffic over networks. Under UNIX/Linux PCAP is implemented through the library *libcap*. The library *WinPcap* is the Windows version of the library *libcap*. Supervision tools can use *pcap* (or *WinPcap*) to capture packets over the network; and to record captured packets in a file and to read saved file.

2) Presentation of the JpCap: Jpcap is an open source library for capturing and sending network packets from Java applications [20]. It provides facilities to:

- capture raw packets live from the wire,
- save captured packets to an offline file, and read captured packets from an offline file,
- automatically identify packet types and generate corresponding Java objects (for Ethernet, IPv4, IPv6, ARP/RARP, TCP, UDP, and ICMPv4 packets),
- filter the packets according to user-specified rules before dispatching them to the application,
- send raw packets to the network.

Jpcap is based on libpcap/winpcap, and is implemented in C and Java programming languages.

Jpcap can be used to develop many kinds of network applications, including network and protocol analyzers, network monitors, traffic loggers, traffic generators, user-level bridges and routers, network intrusion detection systems, network scanners, security tools.

3) Presentation of the JavaMail: The JavaMail API[1] provides classes that model a mail system. JavaMail classes and interfaces are set within four packages namely *javax.mail, java.mail.internet, javax.mail.event, and javax.mail.search*. The javax.mail package defines classes that are common to all mail systems. The *javax.mail.internet* package defines classes that are specific to mail systems based on Internet standards such as MIME, SMTP, POP3, and IMAP. The JavaMail API includes the *javax.mail* package and subpackages.

The JavaMail API is a JDK[2] which is downloadable from the SUN website at the URL *http://java.sun.com/products/javamail*. The JavaMail API is used in this project to alert the system administrator by electronic mail when severe intrusions are detected over the network.

4) Presentation of the MySQL: MySQL [21] is one of the most used database management system over the world. It is used in this work to implement a relational database that stores information about captured packets and generated alarms once an intrusion is detected over the network.

VI. Global Architecture Proposed

Figure 4 presents the global architecture of the proposed network intrusion detection system. It is made of five levels. The first level corresponds to the network listening process and captures packets over this network. At the second level, the packet decoding is done to transmit extracted information to the third level. The intrusion's search in each packet is done at the third level by scanning IP addresses, destinations ports, etc. This information is recorded into a database. At this level, each packet is analyzed to detect a pattern for specific attacks. An alarm is observed when an intrusion pattern is observed. A table of the database records different generated alarms to help an administrator to check the type of attacks. The fourth level corresponds to the main part of the tool. At this level, we implement four dedicated processors for heuristic analysis and a processor to look for patterns. It is possible to implement more or less dedicated processors. The last level is dedicated to the alarms' management and their output mode. In our case, we implement visual alarms and those to be sent by electronic mail in the administrator account.

VII. Implementation and Simulation

The implementation description will take into account the database that stores the captured packets and generated alarms after intrusions' detection.

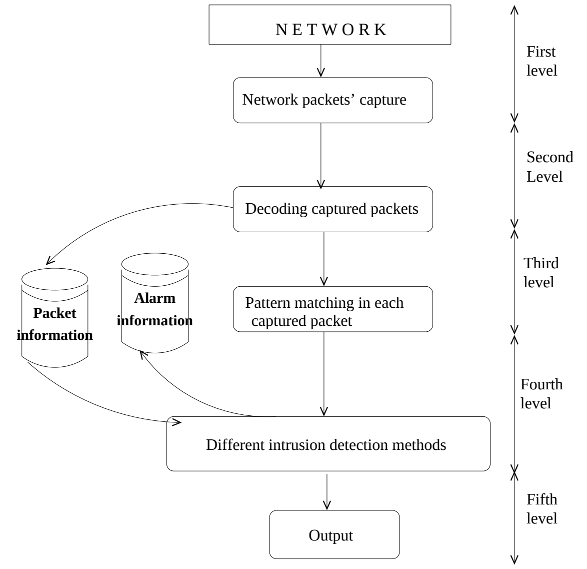


Fig. 4. Global architecture of the proposed network intrusion detection.

A. Description of the Implemented Database

The *MySQL* is used as the relational database management system. The implemented database has four database's tables: Table **TCPCAPTURES** is used to record information about captured TCP packets. Table **UDPCAPTURES** is used to record information about captured UDP packets. Table **ICMPCAPTURES** is used to record information about captured ICMP packets. Finally, the table **DONNEESALERTES** is used to record information about different detected intrusions.

B. Implementation Description

The proposed network intrusion detection system is implemented according to the following five steps, namely listening to the network and capturing the packets, decoding the packets, detecting specific attacks, detecting process heuristically, and printing the output module.

1) Listening to the network and capturing the packets: At this first step, a sniffor is developed using Jpcap library already presented in subsection V-F2. In a Ethernet network, each system has a network card which has its own physical address. The network card examines each packet over the network and catches it once intended to the host machine. One withdraws from this package the various layers such as Ethernet, IP, TCP, etc. to forward information it contains to the application. When a network card is configured in the promiscious mode thanks to the *Jpcap* library, all packets are captured without being out from the traffic.

The sniffer is therefore implemented using the *Jpcap* library through the following steps:

- seeking and printing all network interfaces available on the host machine thanks to the method *JpcapCaptor.getDeviceList()*,
- selecting of the network interface to be used by the sniffer,

[1] Application Programming Interface.
[2] Java Development Kit.

- activating of the network interface onto the proscimous mode thanks to *JpcapCaptor.openDevice()*,
- starting the packets capturing process through the interface *PacketReceiver*

2) Decoding the packets: Packet decoding process also is based on the *Jpcap* library. The decoder receives one after another all the packets from the sniffer and finds their category (TCP, UDP, ICMP, etc.) by comparing them to different available classes in the *Jpcap* library namely *IPPacket, TCPPacket, UDPPAcket, ICMPPacket, etc.* For instance, if the concerned packet is TCP, the decoder collects its source and destination addresses, source and destination ports, data field and TCP flag.

3) Detecting specific attacks: In the proposed architecture, intrusion detection is done at levels 3 and 4. At level 3, a first search of intrusion is done based on the patterns while at level 4 three modules namely *deny of service, Bruteforce, Trojan* based upon heuristic analysis are done.

The heuristic deny of service will serve to detect attacks contained in many packets, which leads to deny of service. There exist numerous attacks of type deny of service. In this work, for the simulation, we are interested in *attacks by land, flood, and death's ping*.

4) Heuristic detection process: Patterns are stored in a database and scanned for intrusion detection.

5) Output module: This module is executed once an attack is detected. It has three distinct modes. The first one is an alarm that informs about intrusion detection. The second mode uses one table in the database for recording attacks through a graphical user interface. The third mode is an alarm through an electronic mail sent to the system administrator. This last mode uses the *Javamail* library.

C. Graphical User Interface

Figure 5 presents the graphical user interface of the developed network intrusion detection system.

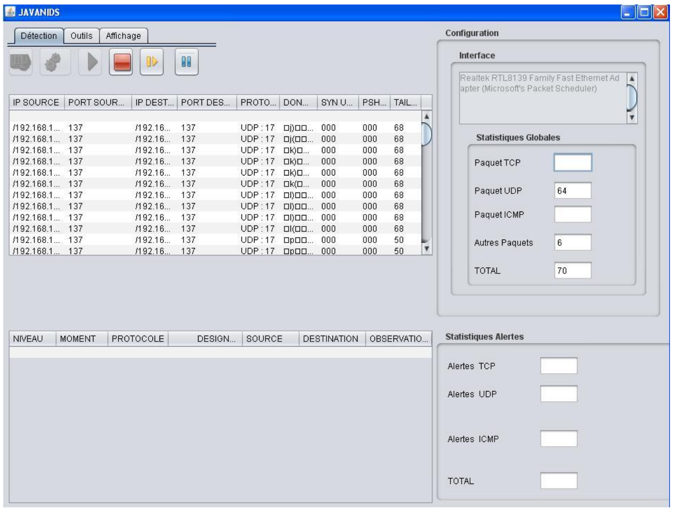
D. Simulation

Our testing methodology is based on simulating computer users - intruders as well as normal users while the intrusion detection system is running. We employed the *hping3* to simulate users in our experiment. Three experiments are carried out to test the proposed network intrusion detection system we installed on a server. The user is simulated by using the *hping* that generates and analyses TCP/IP packets and supports protocols such as TCP, UDP, ICMP, RAW-IP with traceroute mode and many other features [22]. The tool *hping* is installed on one host of the network to simulate different attacks towards other machines in the same network. Three experiments are carried out.

1) First experiment with hping tool by simulating the LAND attack: TCP packets with the same source and destination IP address are sent over the network to simulate the LAND attack through the command

```
# hping3 -n -c 2 -a 192.168.1.123 192.168.1.123
```

Figure 6 presents the behavior of the implemented network intrusion detection system.



IP SOURCE	PORT SOUR...	IP DEST...	PORT DES...	PROTO...	DON...	SYN U...	PSH...	TAIL...
/192.168.1...	137	/192.16...	137	UDP : 17	□j)□□...	000	000	68
/192.168.1...	137	/192.16...	137	UDP : 17	□j(□□...	000	000	68
/192.168.1...	137	/192.16...	137	UDP : 17	□k)□...	000	000	68
/192.168.1...	137	/192.16...	137	UDP : 17	□k)□...	000	000	68
/192.168.1...	137	/192.16...	137	UDP : 17	□k)□...	000	000	68
/192.168.1...	137	/192.16...	137	UDP : 17	□k(□...	000	000	68
/192.168.1...	137	/192.16...	137	UDP : 17	□l)□□...	000	000	68
/192.168.1...	137	/192.16...	137	UDP : 17	□l)□□...	000	000	68
/192.168.1...	137	/192.16...	137	UDP : 17	□l)□□...	000	000	68
/192.168.1...	137	/192.16...	137	UDP : 17	□l(□□...	000	000	68
/192.168.1...	137	/192.16...	137	UDP : 17	□p□□...	000	000	50
/192.168.1...	137	/192.16...	137	UDP : 17	□p□□...	000	000	50

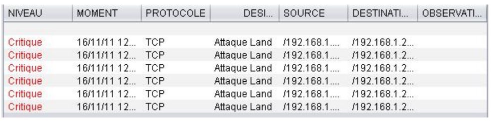
NIVEAU	MOMENT	PROTOCOLE	DESIGN...	SOURCE	DESTINATION	OBSERVATIO...

Realtek RTL8139 Family Fast Ethernet Adapter (Microsoft's Packet Scheduler)

Paquet TCP	
Paquet UDP	64
Paquet ICMP	
Autres Paquets	6
TOTAL	70

Alertes TCP	
Alertes UDP	
Alertes ICMP	
TOTAL	

Fig. 5. Graphical user interface of the proposed network intrusion detection system.



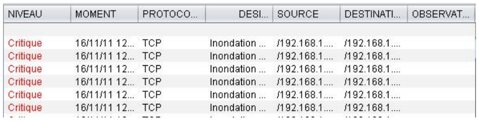
NIVEAU	MOMENT	PROTOCOLE	DESI...	SOURCE	DESTINATI...	OBSERVATI...
Critique	16/11/11 12...	TCP	Attaque Land	/192.168.1....	/192.168.1.2...	
Critique	16/11/11 12...	TCP	Attaque Land	/192.168.1....	/192.168.1.2...	
Critique	16/11/11 12...	TCP	Attaque Land	/192.168.1....	/192.168.1.2...	
Critique	16/11/11 12...	TCP	Attaque Land	/192.168.1....	/192.168.1.2...	
Critique	16/11/11 12...	TCP	Attaque Land	/192.168.1....	/192.168.1.2...	
Critique	16/11/11 12...	TCP	Attaque Land	/192.168.1....	/192.168.1.2...	

Fig. 6. LAND attack detection by the implemented network intrusion detection system.

2) Second experiment with hping tool by simulating flood attack: Flood attacks are simulated towards the host machine with 192.168.1.114 as victim through the command

```
# hping3 -S -p 80 –flood 192.168.1.114
```

Figure 7 presents the behavior of the implemented network intrusion detection system.



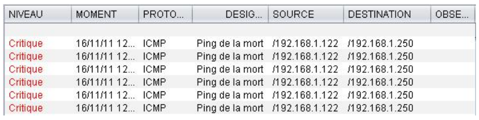
NIVEAU	MOMENT	PROTOCO...	DESI...	SOURCE	DESTINATI...	OBSERVAT...
Critique	16/11/11 12...	TCP	Inondation ...	/192.168.1....	/192.168.1....	
Critique	16/11/11 12...	TCP	Inondation ...	/192.168.1....	/192.168.1....	
Critique	16/11/11 12...	TCP	Inondation ...	/192.168.1....	/192.168.1....	
Critique	16/11/11 12...	TCP	Inondation ...	/192.168.1....	/192.168.1....	
Critique	16/11/11 12...	TCP	Inondation ...	/192.168.1....	/192.168.1....	
Critique	16/11/11 12...	TCP	Inondation ...	/192.168.1....	/192.168.1....	

Fig. 7. Flood attack detection by the implemented network intrusion detection system.

3) Third experiment with hping tool by simulating death's ping attack: Death ping attacks are simulated towards the host machine with 192.168.1.114 as victim through the command

```
# hping3 -l -c 20 192.168.1.114
```

Figure 8 presents the behavior of the implemented network intrusion detection system.



NIVEAU	MOMENT	PROTO...	DESIG...	SOURCE	DESTINATION	OBSE...
Critique	16/11/11 12...	ICMP	Ping de la mort	/192.168.1.122	/192.168.1.250	
Critique	16/11/11 12...	ICMP	Ping de la mort	/192.168.1.122	/192.168.1.250	
Critique	16/11/11 12...	ICMP	Ping de la mort	/192.168.1.122	/192.168.1.250	
Critique	16/11/11 12...	ICMP	Ping de la mort	/192.168.1.122	/192.168.1.250	
Critique	16/11/11 12...	ICMP	Ping de la mort	/192.168.1.122	/192.168.1.250	
Critique	16/11/11 12...	ICMP	Ping de la mort	/192.168.1.122	/192.168.1.250	

Fig. 8. Death's ping attack detection by the implemented network intrusion detection system.

VIII. Conclusion Further Works

In this work, we have proposed an intrusion detection system implemented in Java. This system has been tested by simulating three types of attack: land attack, flooding attack and death ping attack. The proposed system detects all these attacks correctly. The proposed network intrusion detection system is extensible and portable and many other functionalities can be implemented. Nevertheless, it presents some drawbacks. First the proposed system takes into account only the scenario approach. The behavioral approach will be examined in the future.

Evaluating an intrusion detection system is a difficult task. Indeed, it can be difficult even impossible to identify the set of all possible intrusions that might occur at the site where a particular intrusion detection system is employed. To start with, the number of intrusion techniques is quite large [23]. Then, the site may not have access to information about all intrusions that have been detected in the past at other locations. Also, intruders can discover previously unknown vulnerabilities in a computer system, and then use new intrusion techniques to exploit the vulnerablities. Another difficulty in evaluating an intrusion detection system is that although it can ordinary detect a particular intrusion, it may fail to detect some intrusion when the overall level of computing activity in the system is high. This complicates the task of thoroughly testing the intrusion detection system.

In our future work, we will also compare the performance of the proposed network intrusion detection with already existing intrusion detection systems based upon the methodology developed by Puketza [8]. We will also combine the proposed intrusion detection system and the Java-based cryptosystem using a dynamic huffman coding and encryption methods we developed in [24]. So doing, the security is reinforced to avoid intruder to discover plaintext data.

Acknowledgments

We thank anonymous reviewers for their review efforts. We also appreciate comments from our colleagues.

References

[1] J. P. Anderson, "Computer security threat monitoring and surveillance," Fort Washington, Pennsylvania, James P Anderson Co, Tech. Rep., 1980.

[2] D. Denning, "An intrusion-detection model," *IEEE Transaction on Software Engineering*, vol. 13, no. 2, pp. 222–232, 1997.

[3] R. G. Bace, *Intrusion Detection*. Technical Publising, 1995.

[4] B. Mukherjee *et al.*, "Network intrusion detection," *IEEE Network*, vol. 8, no. 3, pp. 26–41, 1994.

[5] K. Ramamohanarao *et al.*, "The curse of ease of access to the internet," *3rd International Conference on Information Systems Security*.

[6] N. Bashah *et al.*, *World Academy of Science, Engineering and Technology*. World Academy of Science, 2005.

[7] K. K. Gupta, "Robust and efficient intrusion detection systems," Ph.D. dissertation, The University of Melbourne, Department of Computer Science and Software Engineering, January 2009.

[8] N. J. Puketza *et al.*, "A methodology for testing intrusion detection systems," *IEEE Transaction on Software Engineering*, vol. 22, no. 10, pp. 719–729, 1996.

[9] M. A. Aydin *et al.*, "A hybrid intrusion detection system design for computer network security," *Computer and Electrical Engineering*, vol. 35, pp. 517–526, 2009.

[10] K. Tan, "The application of neural networks to unix computer security," *IEEE International Conference on Neural Networks*, vol. 1, pp. 476–481, 1995.

[11] H. S. Javitz and A. Valdes, "The sri ides statistical anomaly detector," *IEEE Symposium on Research in Security and Privacy*, pp. 316–376, 1991.

[12] S. Noel *et al.*, *Modern intrusion detection, data mining, and degrees of attack guilt, in applications of data mining in computer security*. Kluwer Academic Publisher, 2002.

[13] N. Debar *et al.*, "A neural network component for an intrusion detection systems," in *IEEE symposium on security and privacy*, 1992, pp. 240–250.

[14] L. M. Gassata, "The artificial immune model for network intrusion detection," in *First international workshop on the recent advances in intrusion detection*, 1998.

[15] J. Kim and P. Bentley, "The artificial immune model for network intrusion detection," in *Seventh European congress on intelligent techniques and soft computing (EUFIT99)*, 1999.

[16] M. Sobirey. (2011, Jan.) Intrusion detection systems. [Online]. Available: http://www-rnks.informatik.tu-cottbus.de/sobirey/ids.html

[17] M. Roesch, "Snort lightweight intrusion detection for networks."

[18] R. Russel, *Snort intrusion detection 2.0*. Rockland, MA: Syngress Publishing, Inc, 2003.

[19] D. Burgermeister and J. Krier. (2010, Dec.) Système de détection d'intrusion. [Online]. Available: http://www.dbprog.developpez.com/securite/ids/IDS.pdf

[20] K. Fujii. (2007, Jan.) Jpcap tutorial. [Online]. Available: http://netresearch.ics.uci.edu/kfujii/Jpcap/doc/tutorial/index.html

[21] C. Thibaud, *MySQL 5: installation, mise en oeuvre, administration et programmation*. Edition Eyrolles, 2006.

[22] N. Cheswick and S. Bellovin, *Firewalls and Internet Security: Repelling the Willy Hacker*. Pearson Education Inc., 2003.

[23] P. G. Neumann and D. Parker, "A summary of computer misuse techniques," in *12th National Computer Security Conference, Baltimore, MD*, 1989, pp. 396–407.

[24] E. C. Ezin, "Implementation in java of a cryptosystem using a dynamic huffman coding and encryption methods," *International Journal of Computer Science and Information Security*, vol. 9, no. 3, pp. 154–159, 2011.

Authors' profiles

Eugène C. Ezin received his Ph.D degree with highest level of distinction in 2001 after research works carried out on neural and fuzzy systems for speech applications at the International Institute for Advanced Scientific Studies in Italy. Since 2007, he has been a senior lecturer in computer science. He is a reviewer of Mexican International Conference on Artificial Intelligence. His research interests include neural network and fuzzy systems, high performance computing, signal processing, cryptography, modeling and simulation.

Hervé Guy Akakpo received his MSc in computer science with highest level of distinction in 2011. He is currently employed at the Caisse Autonome d'Amortissement. and affiliated to the Institut de Mathématiques et de Sciences Physiques within the master program of computer science for tutoring the course on networking. His research interests include information system and network security.

Using Biometric techniques to secure online student assessment: comparative study

Jamaludin Ibrahim
Faculty of Information and Communication Technology
IIUM
Kuala Lumpur, Malaysia
jamaludinibrahim@iium.edu.my

Muna A. Ali, Rasheed Nassr
Faculty of Information and Communication Technology
IIUM
Kuala Lumpur, Malaysia

Abstract— **Currently E-learning systems do not provide a tool to authenticate student continuously during online assessment, this raises the probability of cheating. Many proposed solutions use different biometric techniques to identify and authenticate student continuously, they use different techniques with different measures. This paper proposes certain criteria that should be available in any proposed biometric technique to be fitted with e-learning architecture and continues authentication of student during online assessment. This paper investigates some proposed solutions to see compatibility of those solutions with the proposed criteria.**

Keywords-component; Biometric, E-learning, online assessment

I. Introduction

Many educational organizations depend on E-learning system to conduct education; this dependency was focusing on delivering material online and facilitating interaction among students and instructors but now increases and reaches to the level of issuing trustful certificates as a result of that the need for confidential and trustful security mechanism is highly required. Most of current e-learning systems pay less attention for continues online assessment security and focus on securing information assets [2, 1, and 10]. E-learning system might fail to guarantee the real identity of the remote user and whether does the intended student who is doing the assessment or somebody on behalf during assessment session [5, 1, 6, and 12]. Current and future requirements of E-learning system may require more security that use physical or/and behavioral features of the learners, further details about the need for biometric security techniques in E-learning can be found in [12].

II. Literature review

Even though Ref [3] deeply analyzed the challenges and opportunities of E-learning and visualized the shape of future E-learning, security challenges were not mentioned or analyzed; however Ref [2] reported the importance of security for E-learning; the focus was how to protect data from unauthorized access. As a matter of fact, E-learning system just resembles any other systems, needs to protect its data but the difference is the need to identify and verify student's identity continuously during assessment session. Apart, Biometric devices' prices are getting decrease and various biometric devices currently can be found embedded in most of laptops; this trend may encourage E-learning system designers to consider biometric authentication in future design [8, and 12]. Besides that, the lack of secured and granted students' identification system in e-learning online assessment could limit the success and extendibility of E-learning system [9]. Usually biometric used to control access to restricted physical environment and used rarely to authenticate remote users. Traditional authentication techniques such as password do not prevent student's larceny and are transferable from user to another [8, Chang 2009; 5], because of failing to grant student's identity by using password, it has been suggested to restrict accessing online assessment from certain areas [8]. However that cannot be practical solution particularly when e-learning system's boundaries exceed campus area as well as it requires students to be locked in certain area for exam period which may be difficult to fulfill.

Biometric techniques may be evaluated according to acceptance, features, durability, universality, and permanence [10, 12, and 13]. Many literatures have discussed the flexibility, acceptance, performance and strong and weak points of most of the biometric techniques [10, 12, and 13]. However; it is not necessary to be fully considered in e-learning because other features may control choosing the suitable biometric for E-learning such as ability of identification process to break down into client and server procedures, required equipments, cost-effective and does it affect E-learning performance. In case of required equipment, Chang (2009) proposal-personalized rhythm- to authenticate users does not require special biometric devices to verify student's identity but detects student's identity by the way she/he behaves with input devices- mouse, keyboard, etc. and neural network is used to decrease a chances of false acceptance rate. Though Face recognition is not accurate as much as fingerprint [11], it seems the most eligible because it can prove the physical existence of the remote user in front of camera as well as availability of camera in most of the current manufactured laptops strengthens its eligibility. Reference [11] described the process of Face recognition for remote user by the following steps: a- extract the image portion that contains the face, b- consider other factors such as distance to make the extracted part of image comparable, c- search for matching in database. Face recognition is influenced by pose and illumination; it needs extra transformation techniques to decrease the effects of such problems[11]. Similarly; video face recognition can use the same algorithms of face recognition from image, the only difference is processing more than one image, however; some algorithms use the unique features of video such as continuity to recognize face, while other techniques use image and video together to recognize face [11]**.** There is a little superiority for video against image but still has own problems such as quality, variations of illumination and facial expressions [11]. To increase the possibility of face recognition, pictures are taken in different poses: reading, looking, and typing of each identity in order to be used during verification [11].

III. Candidate Biometric method's features

Beside the features that already mentioned in section I, biometric methods needs extra features that qualify it to fit E-learning architecture as well as the ability to sustain continues authentication during online assessment. Biometric method may need to meet the following: divisible into client/server procedures to work with client/server architecture of E-learning system, ability to use partial features to identify person- to make it fast and real time, law network bandwidth consuming because E-learning's nature that may utilize multimedia in education that already consumes network bandwidth and the candidate method should not make it worse, eventually; capability to identify real personality of the examinee. This section will investigate some biometric techniques that are proposed to see the compatibility with those extra features.

The current biometric techniques that can be run without extra equipments are keystroke pattern, face and facial, and voice recognition because most of the current computing devices equipped by camera and microphone. Apart; It is known that voice recognition process may be effected by human health conditions and it is uncomfortable for student to speak from time to time to confirm his existence, as it is possible the person who conduct online assessment is attend the session with somebody else where student keep authenticate himself while the second one doing assessment on behalf, voice recognition is not recommended because noise and replay attack [13]. Those circumstances may exclude utilizing voice recognition as continues authentication technique. However; Ref [11] face recognition-based model fits client/server architecture of E-learning as well as it does not send the face image as a whole but certain vectors that capture the main biometric features. This will decrease the time to send data through Internet and time to process the whole image features. In contrast; COPaCII model has capability to identify face facial even with law resolutions images but all work done in the server-side, it lacks client/ server architecture of E-learning [6]. While Ref [4] established a model to evaluate the performance of hand geometric and face recognition without need for large biometric feature, that model proves that there is no need for large number of features to identify person. Besides that the main concern of that proposed model was the performance of the biometric system. This model seems to be suitable because it meet most of the features but the architecture of this model was not elaborated but it could be expected that it will be easier to integrate it with E-learning particularly face-recognition's part that supports the idea of continues online assessment. Reference [13] gave estimated minimum and maximum memory space that could be consumed of some biometric, table I shows their memory size estimation, and this paper adds factor of ability to be applied to identify real personality and existance of the examinee during online assessment. Facial recognition has various memory sizes depend on whether full or partial features will be collected and it has high ability to ensure real personality existence in front of camera. While Iris is more accurate but consumes somehow high memory space, the problem with Iris is that it is difficult to propose practical solution using this method because it is not reasonable to convince student to Stare at camera all the time during online assessment.

Reference [7] reported that the performance of the current face recognition systems is reasonable, and concluded that it is questionable to verify identity by face recognition technique only without help of certain contextual information. But it looks as the most suitable tool to be included in e-learning system. Similarly; keystroke rhythm does not interrupt students and the same time detects availability of the students during online assessment [10]. A detailed list of considerations that must be bear in mind when biometric techniques are chosen to be implemented in E-learning can be found in Ref [10]. Reference [4] addressed how to evaluate performance of several biometric techniques, it is found that biometric characteristics could be normally distributed particularly hand geometric and human face. However; the main considerable key factors to evaluate biometric system are False Acceptance Rate (*FAR) and* False Rejection Rate (FRR) [4].

IV. Conclusion

It is practical to include biometric authentication methods with E-learning however; some considerations must be bear in mind such as ability to identify personality of examinee in real-time and ability to be divided into client/server procedures. Some biometric does not cost too much such as keystroke method and it is practical, while others need some equipments that already become common nowadays meanwhile voice and iris could not be practical, because the former one has security, technical and human-related problems; later consumes more memory and has adaptability problem with students. It is expected that biometric methods be included in current and future E-learning system as a result of decreasing prices of equipments and availability of technical and programming tools to utilize biometric methods to authenticate E-learning online assessment.

Biometric Method	Estimated Memory size	Real Personality existence
Retina	96B- 10KB	
Facial recognition	96 bytes to 5 KB	Yes
Hand geometric	9 B	No
Iris	about 256B	Yes
Voice recognition	….	No
Keystroke	….	yes

Table I. Memory size estimated and ability to identify student during online assessment

REFERENCES

[1] E. G. Agulla, L. A. Rifón, J. L.Castro, and C. G. Mateo, "Is my student at the other side? Applying Biometric Web Authentication to e-learning. Eighth International Conference on Advanced Learning Technologies" IEEE, pp. 551-55, 2008.

[2] S. R. Balasundaram , "Securing Tests in E-Learning Environment. ICCCS'11" ACM, Rourkela, Odisha, India, pp. 624-627, 2011.

[3] V. Cantoni, M. Cellario, and M. Porta, Perspectives and challenges in e-learning: towards natural interaction paradigms. Visual Languages and Computing , 333–345, 2004.

[4] M. Golfarelli, D. Maio, and D. Maltoni,. On the Error-Reject Trade-Off in Biometric Verification Systems. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE , 786-796, 1997.

[5] A. J. Harris, and D. C. Yen, biometric authentication: assuring access to information. Information management & computer security , 12 – 19, 2002.

[6] R. INABA, E. WATANABE, and K. KODATE, Security ApplicationS Of Optical Face Recognition System: Access Control in E-Learning. OPTICAL REVIEW , 255- 261,2003.

[7] A. Jain, A. Ross, and S. Prabhakar, An Introduction to Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology , 4-20, 2004.

[8] E. Marais, D. Argles, and B. V. Solms, "Security issues specific to E-assessments. 8th Annual Conference on WWW Applications ". Bloemfontein: The ECS EPrints.

[9] A. Marcus, J. Raul, R. Ramirez-Velarde, and J. Nolazco-Flores, Addressing Secure Assessments for Internet-based Distance Learning: Still an unresolvable issue? Niee, 2008 .

[10] A. Moini, and A. M. Madni, Leveraging Biometrics for User Authentication in Online Learning: A Systems Perspective. IEEE SYSTEMS JOURNAL , 469-476, 2009.

[11] B. E. Penteado, and A. N. Marana, A VIDEO-BASED BIOMETRIC AUTHENTICATION FOR ELEARNING WEB APPLICATIONS. In *Lecture Notes in Business Information Processing*, springerlink, 770-779, 2009.

[12] K. Rabuzin, M. Bac'a, and M. Sajko," E-learning: Biometrics as a Security Factor. Proceedings of the International Multi-Conference on Computing in the Global Information Technology ICCGI'06" IEEE, pp. 64, 2006.

[13] V. Zorkadis, and P. Donos, On biometrics-based authentication and identification from a privacy-protection perspective Deriving privacy-enhancing requirements, Information Management & Computer Security, pp. 125-137, 2004.

Training of Feed-Forward Neural Networks for Pattern-Classification Applications Using Music Inspired Algorithm

Ali Kattan
School of Computer Science,
Universiti Sains Malaysia,
Penang 11800, Malaysia
kattan@cs.usm.my

Rosni Abdullah
School of Computer Science,
Universiti Sains Malaysia,
Penang 11800, Malaysia
rosni@cs.usm.my

Abstract—**There have been numerous biologically inspired algorithms used to train feed-forward artificial neural networks such as generic algorithms, particle swarm optimization and ant colony optimization. The Harmony Search (HS) algorithm is a stochastic meta-heuristic that is inspired from the improvisation process of musicians. HS is used as an optimization method and reported to be a competitive alternative. This paper proposes two novel HS-based supervised training methods for feed-forward neural networks. Using a set of pattern-classification problems, the proposed methods are verified against other common methods. Results indicate that the proposed methods are on par or better in terms of overall recognition accuracy and convergence time.**

Keywords-harmony search; evolutionary methods; feed-forward neural networks; supervised training; pattern-classification

I. INTRODUCTION

Harmony Search (HS) is a relatively young meta-heuristic stochastic global optimization (SGO) method [1]. HS is similar in concept to other SGO methods such as genetic algorithms (GA), particle swarm optimization (PSO) and ant colony optimization (ACO) in terms of combining the rules of randomness to imitate the process that inspired it. However, HS draws its inspiration not from biological or physical processes but from the improvisation process of musicians. HS have been used successfully in many engineering and scientific applications achieving better or on par results in comparison with other SGO methods [2-6]. HS is being compared against other evolutionary based methods such as GA where a significant amount of research has already been carried out on the application of HS in solving various optimization problems [7-11]. The search mechanism of HS has been explained analytically within a statistical-mathematical framework [12, 13] and was found to be good at identifying the high performance regions of solution space within reasonable amount of time [14]. Enhanced versions of HS have been proposed such as the Improved Harmony Search (IHS) [15] and the Global-best Harmony Search (GHS) [16], where better results have been achieved in comparison with the original HS when applied on some integer programming problems. HS, IHS variants are being used in many recent works [17].

Evolutionary based supervised training of feed-forward artificial neural networks (FFANN) using SGO methods, such as GA, PSO and ACO has been already addressed in the literature [18-25]. The authors have already published a method for training FFANN for a binary classification problem (Cancer) [27] which has been cited in some recent works [28]. This work is an expanded version of the original work that includes additional classification problems and a more in depth discussion and analysis. In addition to the training method published in [27] this work presents the adaptation for the original IHS optimization method [15]. Then IHS is modified to produce the second method using a new criterion, namely the best-to-worst (BtW) ratio, instead of the improvisation count for determining the values of IHS's dynamic probabilistic parameters as well as the termination condition. Implementation considers pattern-classification benchmarking problems to compare the proposed techniques against GA-based training as well as the standard Backpropagation (BP) training.

The rest of this work is organized as follows. Section II presents a literature review of related work; Section III introduces the HS algorithm, its parameters and modeling; section IV introduces the IHS algorithm indicating the main differences from the original HS; section V introduces the proposed methods discussing the adaptation process in terms of FFANN data structure, HS memory remodeling and fitness function introducing a complete training algorithm and the initial parameters settings; section VI covers the results and discussion. Conclusions are finally made in section VII.

II. RELATED WORK

The supervised training of an artificial neural network (ANN) involves a repetitive process of presenting a training data set to the network's input and determining the error between the actual network's output and the intended target output. The individual neuron weights are then adjusted to minimize such error and give the ANN its generalization ability. This iterative process continues until some termination condition is satisfied. This usually happens based on some measure, calculated or estimated, indicating that the current achieved solution is presumably good enough to stop training [29]. FFANN is a type of ANNs that is characterized by a topology with no closed paths and no lateral connections existing between the neurons in a given layer or back to a previous one [29]. A neuron in a given layer is fully connected to all neurons in the subsequent layer. The training process of FFANNs could also involve the network's structure represented by the number of hidden layers and the number of neurons within each [30-32]. FFANNs having a topology of just a single hidden layer, which sometimes referred to as 3-layer FFANNs, are considered as universal approximators for arbitrary finite-input environment measures [33-36]. Such configuration has proved its ability to match very complex patterns due to its capability to learn by example using relatively simple set of computations [37]. FFANNs used for pattern-classification have more than one output unit in its output layer to designate "classes" or "groups" belonging to a certain type [34, 38]. The unit that produces the highest output among other units would indicate the winning class, a technique that is known the "winner-take-all" [39, 40].

One of the most popular supervised training methods for FFANN is the BP learning [36, 41, 42]. BP is basically a trajectory-driven method, which is analogous to an error-minimizing process requiring that the neuron transfer function to be differentiable. The concept is illustrated in Fig. 1 using a 3-dimensional error surface where the gradient is used to locate minima points and the information is used to adjust the network weights accordingly in order to minimize the output error [43].

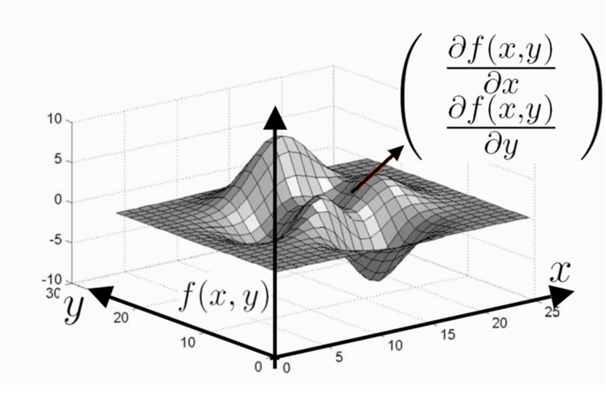


Figure 1. An illustration of the gradient-descent technique using a 3-dimensional error surface

However, BP is generally considered to be inefficient in searching for global minimum of the search space [44] since the BP training process is associated with two major problems; slow convergence for complex problems and local minima entrapment [36, 45]. ANNs tend to generate complex error surfaces with multiple local minima and trajectory-driven methods such as BP possess the possibility of being trapped in local solution that is not global [46]. Different techniques have been proposed to cure these problems to a certain extent including techniques such as simulated annealing and dynamic tunneling [36] as well as using special weight initialization techniques such as the Nguyen-Widrow method [39, 47, 48]. BP could also use a momentum constant in it's learning rule, where such technique accelerates the training process in flat regions of the error surface and prevents fluctuations in the weights [42].

Evolutionary supervised training methods offer an alternative to trajectory-driven methods. These are SGO techniques that are the result of combining an evolutionary optimization algorithm with the ANN learning process [31]. Evolutionary optimization algorithms are usually inspired form biological processes such as GA [44], ACO [49], Improved Bacterial Chemo-taxis Optimization (IBCO) [50], and PSO [51]. Such evolutionary methods are expected to avoid local minima frequently by promoting exploration of the search space. Their explorative search features differ from those of BP in that they are not trajectory-driven, but population driven. Using an evolutionary ANN supervised training model would involve using a fitness function where several types of these have been used. Common fitness functions include the ANN sum of squared errors (SSE) [20, 52, 53], the ANN mean squared error (MSE) [49-51], the ANN Squared Error Percentage (SEP) and the Classification Error Percentage (CEP) [18, 54]. The common factor between all of these forms of fitness functions is the use of ANN output error term where the goal is usually to minimize such error. Trajectory-driven methods such as BP have also used SSE, among others, as a training criterion [39, 43].

Many evolutionary-based training techniques have also reported to be superior in comparison with the BP technique [44, 49, 50, 54]. However, most of these reported improvements were based on using the classical XOR ANN problem. It was proven that the XOR problem has no local minima [55]. In addition, the size of the training data set of this problem is too small to generalize the superiority of any training method against others.

III. The Harmony Search Algorithm

The process of music improvisation takes place when each musician in a band tests and plays a note on his instrument. An aesthetic quality measure would determine if the resultant tones are considered to be in harmony with the rest of the band. Such improvisation process is mostly noted in Jazz music where the challenge is to make the rhythm section sound as cool and varied as possible without losing the underlying groove [56].

Each instrument would have a permissible range of notes that can be played representing the pitch value range of that musical instrument. Each musician has three basic ways to improvise a new harmony. The musician would either play a totally new random note from the permissible range of notes, play an existing note from memory, or play a note from memory that is slightly modified. Musicians would keep and remember only good improvised harmonies till better ones are found and replace the worst ones.

The basic HS algorithm proposed by Lee and Geem [57] and referred to as "classical" [13] uses the above scenario as an analogy where note played by a musician represents one component of the solution vector of all musician notes and as shown in Fig. 2.

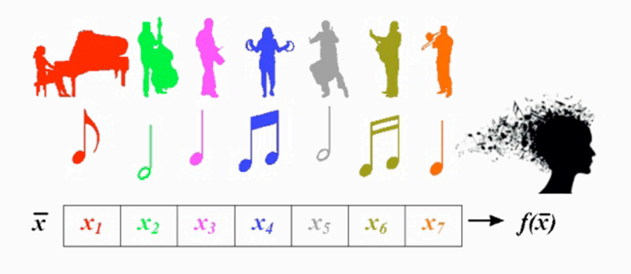


Figure 2. Music improvisation process for a harmony in a band of seven

The best solution vector is found when each component value is optimal based on some objective function evaluated for this solution vector [3]. The number of components in each vector N represents the total number of decision variables and is analogous to the tone's pitch, i.e. note values played by N musical instruments. Each pitch value is drawn from a pre-specified range of values representing the permissible pitch range of that instrument. A Harmony Memory (HM) is a matrix of the best solution vectors attained so far. The HM size (HMS) is set prior to running the algorithm. The ranges' lower and upper limits are specified by two vectors $\mathbf{x}^L$ and $\mathbf{x}^U$ both having the same length N. Each harmony vector is also associated with a harmony quality value (fitness) based on an objective function ***f(x)***. Fig. 3 shows the modeling of HM.

Improvising a new-harmony vector would consider each decision variable separately where HS uses certain parameters to reflect playing probabilistic choices. These are the Harmony Memory Considering Rate (HMCR) and the Pitch Adjustment Rate (PAR). The former determines the probability of playing a pitch from memory or playing a totally new random one. The second, PAR, determines the probability of whether the pitch that is played from memory is to be adjusted or not. Adjustment value for each decision variable is drawn from the respective component in the bandwidth vector **B** having the size N.

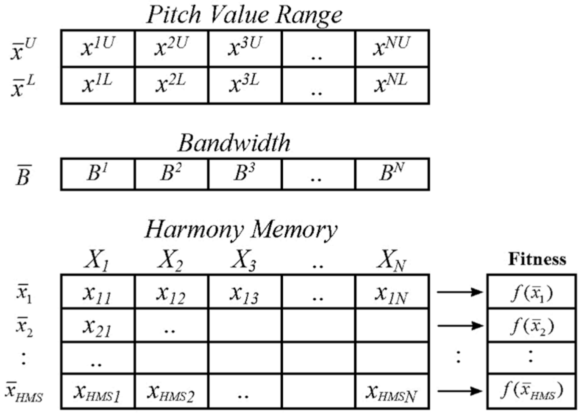


Figure 3. The modeling of HM with N decision variables

The adjustment process should guarantee that the resultant pitch value is within the permissible range specified by $\mathbf{x}^L$ and $\mathbf{x}^U$. The classical HS algorithm pseudo code is given in Algorithm 1.

```
1  Initialize the algorithm parameters (HMS, HMCR, PAR, B, MAXIMP)
2  Initialize the harmony memory HM with random values drawn from vectors [x^L,x^U]
3  Iteration itr=0
4  While itr < MAXIMP Do
5      Improvise new harmony vector x'
6              Harmony Memory Considering:
                   $x_i' \leftarrow \begin{cases} x_i' \in \{x_{i1}, x_{i2}, ..., x_{iHMS}\} & \text{with probability HMCR} \\ x_i' \in X_i & \text{with probability (1-HMCR)} \end{cases}$
               If probability HMCR Then
                 Pitch adjusting:
7                  $x_i' \leftarrow \begin{cases} x_i' \pm rand(0,1) \cdot B_i & \text{with probability PAR} \\ x_i' & \text{with probability (1-PAR)} \end{cases}$
                 Bounds check:
                   $x_i' \leftarrow \min(\max(x_i', x_i^L), x_i^U)$
               EndIf
8      If x' is better than the worst harmony in HM Then Replace worst harmony in HM with x'
9      itr= itr+1
10 EndWhile
11 Best harmony vector in HM is the solution
```

Algorithm 1. Pseudo code for the classical HS algorithm

The improvisation process is repeated iteratively until a maximum number of improvisations MAXIMP is reached. Termination in HS is determined solely by the value of MAXIMP. The choice of this value is a subjective issue and has nothing to do with the quality of the best-attained solution [16, 58, 59].

The use of solution vectors stored in HM is similar to the genetic pool in GA in generating offspring based on past information [10]. However, HS generates a new solution vector utilizing all current HM vectors not just two (parents) as in GA. In addition, HS would consider each decision variable independently without the need to preserve the structure of the gene.

IV. The Improved Harmony Search Algorithm

Mahdavi et al. [15] have proposed the IHS algorithm for better fine-tuning of the final solution in comparison with the classical HS algorithm. The main difference between IHS and the classical HS is that the two probabilistic parameters namely PAR and B, are not set statically before run-time rather than being adjusted dynamically during run-time as a function of the current improvisation count, i.e. iteration (*itr*), bounded by MAXIMP. PAR would be adjusted in a linear fashion as given in equation (1) and shown in Fig. 4(a). B on the other hand would decrease exponentially as given in equation (2) and (3) and shown in Fig. 4(b). Referring to classical HS given in Algorithm 1, this adjustment process takes place just before improvising new harmony vector (line 5). PAR_{min} and PAR_{max} would replace the initial parameter PAR and B_{max} and B_{min} would replace the initial parameter B (line 1).

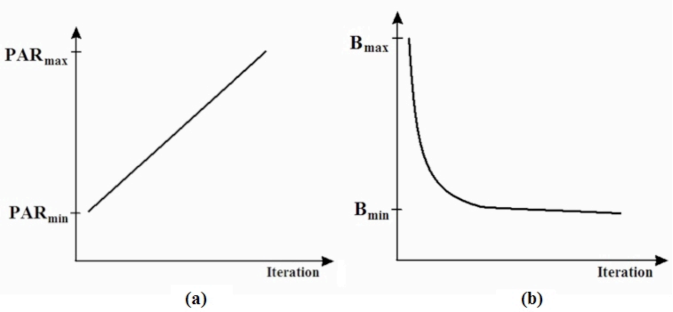


Figure 4. The adjustment of the probablistic parameters in IHS
(a) dynamic PAR value increases linearly as a function of iteration number,
(b) dynamic B value decreases exponentially as a function of iteration number

$$PAR(itr) = PAR_{\min} + \frac{PAR_{\max} - PAR_{\min}}{MAXIMP} \times itr \tag{1}$$

$$B(itr) = B_{\max} \exp(c \cdot itr) \tag{2}$$

$$c = \ln(\frac{B_{\min}}{B_{\max}}) / MAXIMP \tag{3}$$

PAR, which determines if the value selected from HM is to be adjusted or not, starts at PAR$_{\min}$ and increases linearly as a function of the current iteration count with a maximum limit at PAR$_{\max}$. So as the iteration count becomes close to MAXIMP, pitch adjusting would have a higher probability. On the other hand B, the bandwidth, starts high at B$_{\max}$ and decreases exponentially as a function of the current iteration count with a minimum limit at B$_{\min}$. B tends to be smaller in value as the iteration count reaches MAXIMP allowing smaller changes.

V. Proposed Methods

The proposed supervised FFANN training method considers the aforementioned IHS algorithm suggested in [15]. In order to adapt IHS for such a task, suitable FFANN data structure, fitness function, and training termination condition must be devised. In addition, the HM must be remodeled to suit the FFANN training process. Each of these is considered in the following sections.

A. FFANN data structure

Real-coded weight representation was used in GA-based ANN training methods, where such technique proved to be more efficient in comparison with the binary-coded one [52, 53]. It has been shown that binary representation is neither necessary nor beneficial and it limits the effectiveness of GA [46]. The vector representation from the Genetic Adaptive Neural Network Training (GANNT) algorithm originally introduced by Dorsey et al. [18, 20, 53, 60] was adopted for the proposed method. Fig. 5 illustrates such representation for a small-scale sample FFANN. Each vector represents a complete set of FFANN weights including biases where weight values are treated as genes. Neurons respective weights are listed in sequence assuming a fixed FFANN structure.

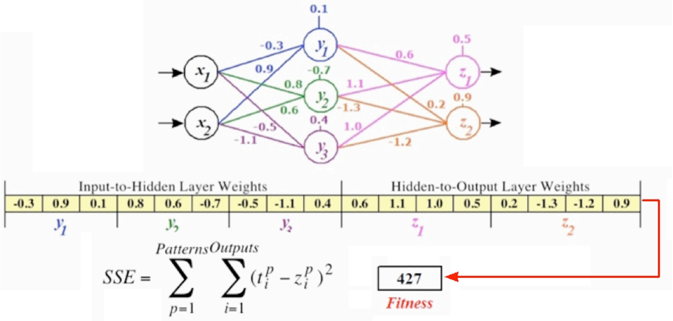


Figure 5. Harmony vector representation of FFANN weights

B. HM remodeling

Since FFANN weight values are usually within the same range, the adapted IHS model could be simplified by using fixed ranges for all decision variables instead of the vectors $\mathbf{x}^{\mathbf{L}}$, $\mathbf{x}^{\mathbf{U}}$ and $\mathbf{B}$. This is analogous to having the same musical instrument for each of the N decision variables. Thus the scalar range [x^{L}, x^{U}] would replace the vectors $\mathbf{x}^{\mathbf{L}}$, $\mathbf{x}^{\mathbf{U}}$ and the scalar value B would replace the vector $\mathbf{B}$. The B value specifies the range of permissible weight changes given by the range [-B,B]. The remodeled version of HM is shown in Fig. 6 with one "Fitness" column. If the problem considered uses more than one fitness measure then more columns are added.

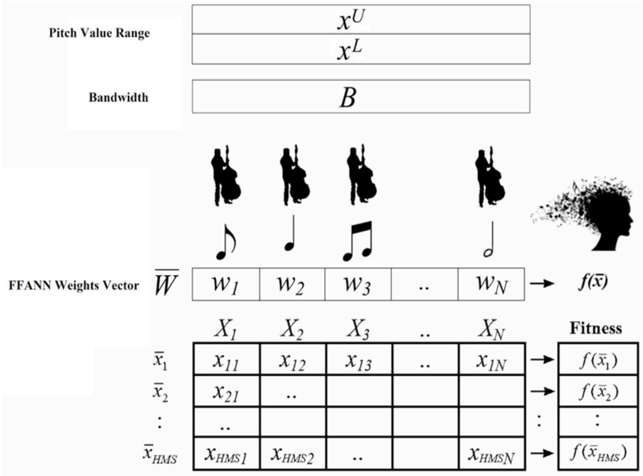


Figure 6. Adapted HM model for FFANN training

C. Fitness function & HS-based training

The proposed method uses SSE as its main fitness function where the goal is to minimize the amount of this error [43]. SSE is the squared difference between the target output and actual output and this error is represented as $(t-z)^2$ for each pattern and each output unit and as shown in Fig. 5. Calculating SSE would involve doing FFANN forward-pass calculations to compare the resultant output with target output. Equations (4) through (6) give these calculations assuming a bipolar sigmoid neuron transfer function [39].

Considering the use of FFANNs for pattern classification networks, CEP, given in (7), could be used to complement SSE's raw error values since it reports in a high-level manner the quality of the trained network [54].

$$SSE = \sum_{p=1}^{P}\sum_{i=1}^{S}(t_i^p - z_i^p)^2 \qquad (4)$$

$$y = \sum_{i=1}^{n+1} w_i x_i \qquad (5)$$

$$z = F(y) = \frac{2}{1+e^{-y}} - 1 \qquad (6)$$

$$CEP = \frac{E_P}{P} \cdot 100\% \qquad (7)$$

where

P	total number of training patterns
S	total number of output units (classes)
t	target output (unit)
z	actual neuron output (unit)
y	sum of the neuron's input signals
w_i	the weight between this neuron and unit i of previous layer (wn+1 represents bias)
x_i	input value from unit I of previous layer (output of that unit)
$n+1$	total number of input connections including bias
$F(y)$	neuron transfer function (bipolar sigmoid)
E_p	total number of incorrectly recognized training patterns

The flowchart shown in Fig. 7 presents a generic HS-based FFANN training approach that utilizes the HM model introduced above. The algorithm would start by initializing the HM with random harmony vectors representing candidate FFANN weight vector values. A separate module representing the problem's FFANN computes each vector's fitness individually. This occurs by loading the weight vector into the FFANN structure first then computing the fitness measure, such as SSE and CEP, for the problem's data set by performing forward-pass computations for each training pattern. Then each vector is stored in HM along with its fitness value(s). An HM fitness measure could be computed upon completing the initialization process. Such measure would take into considerations all the HM fitness values stored such as an average fitness. The training would then proceed in a similar fashion to Algorithm 1 by improvising new weight vector, finding its fitness and deciding whether to insert in HM or not. The shaded flowchart parts in Fig. 7 are to be customized by each of the IHS-based proposed training methods introduced next.

D. *The adapted IHS-based training algorithm*

The IHS algorithm is adapted to use the data structure and the remodeled HM introduced above. The newly improvised harmony is accepted if its SSE value is less than that of the worst in HM and its CEP value is less than or equal to the average value of CEP in HM. The latter condition would guarantee that the newly accepted harmonies would yield the same or better overall recognition percentage. The justification can be explained by considering the winner-take-all approach used for the pattern-classification problems considered. Lower CEP values are not necessarily associated with lower SSE values. This stems from the fact that even if the SSE value is small, it is the winner class, i.e. the one with the highest value, is what determines the result of the classification process.

Fig. 8 shows the flowchart of the adapted IHS training algorithm, which is a customized version of the one given earlier in Fig. 7. Improvising a new harmony vector, which is a new set of FFANN weights, is given as pseudo code of Algorithm 2.

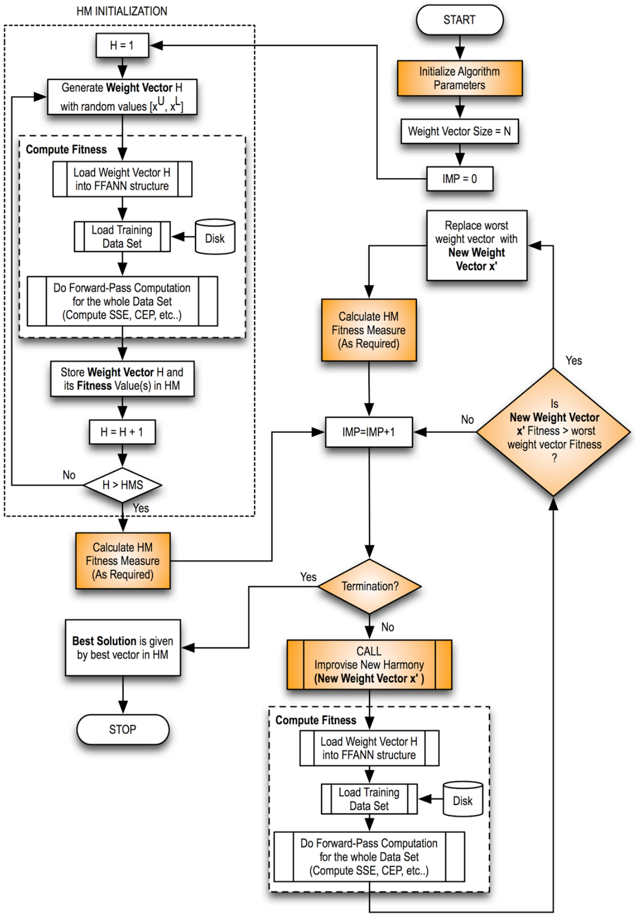

HM INITIALIZATION
H = 1
Generate Weight Vector H with random values [x^U, x^L]
Compute Fitness
Load Weight Vector H into FFANN structure
Load Training Data Set
Disk
Do Forward-Pass Computation for the whole Data Set (Compute SSE, CEP, etc..)
Store Weight Vector H and its Fitness Value(s) in HM
H = H + 1
No
H > HMS
Yes
Calculate HM Fitness Measure (As Required)
Best Solution is given by best vector in HM
STOP
START
Initialize Algorithm Parameters
Weight Vector Size = N
IMP = 0
Replace worst weight vector with New Weight Vector x'
Calculate HM Fitness Measure (As Required)
Yes
No
Is New Weight Vector x' Fitness > worst weight vector Fitness ?
IMP=IMP+1
Yes
Termination?
No
CALL Improvise New Harmony (New Weight Vector x')
Compute Fitness
Load Weight Vector H into FFANN structure
Load Training Data Set
Disk
Do Forward-Pass Computation for the whole Data Set (Compute SSE, CEP, etc..)

Figure 7. Generic FFANN training using adapted HS-based algorithm

E. *The modified IHS-based training algorithm using BtW ratio*

In the plain version of the adapted IHS training algorithm discussed in the previous section, MAXIMP value would affect the rate of change for PAR and B as well as being the only termination condition of the algorithm. Selecting a value for

MAXIMP is a subjective issue that is merely used to indicate the total number of times the improvisation process is to be repeated. The modified version of IHS uses a quality measure of HM represented by the BtW criterion. BtW is a new parameter representing the ratio of the current best harmony fitness to the current worst harmony fitness in HM. With SSE taken as the main fitness function, the BtW value is given by the ratio of the current best harmony SSE to the current worst harmony SSE and as given in equation (8).

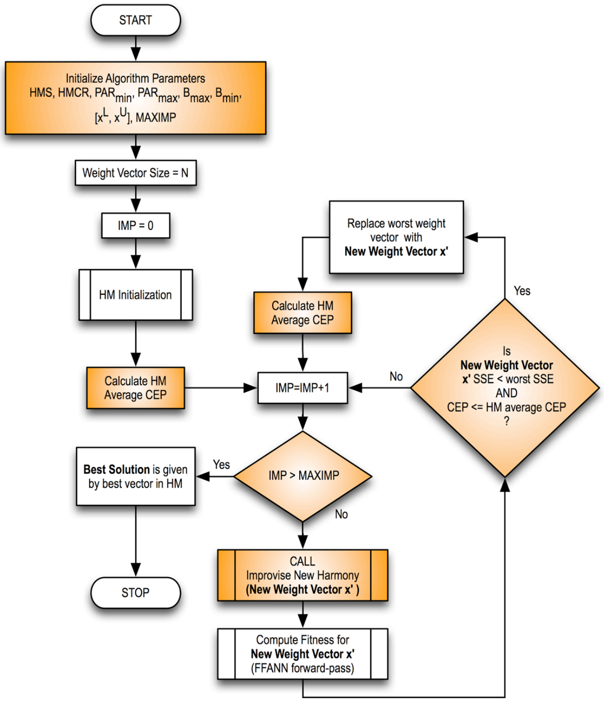

START
Initialize Algorithm Parameters
HMS, HMCR, PARmin, PARmax, Bmax, Bmin,
[xL, xU], MAXIMP
Weight Vector Size = N
IMP = 0
HM Initialization
Calculate HM Average CEP
Replace worst weight vector with New Weight Vector x'
Calculate HM Average CEP
Yes
Is New Weight Vector x' SSE < worst SSE AND CEP <= HM average CEP ?
No
IMP=IMP+1
Best Solution is given by best vector in HM
Yes
IMP > MAXIMP
No
STOP
CALL Improvise New Harmony (New Weight Vector x')
Compute Fitness for New Weight Vector x' (FFANN forward-pass)

Fig. 8. FFANN training using adapted IHS algorithm

```
Create new harmony vector x' of size N
For i=0 to N do
  RND= Random(0,1)
  If (RND<=HMCR) //harmony memory considering
      RND= Random(0,HMS)
      x'(i)= HM(RND,i) //harmony memory access
       PAR= $PAR_{min}$+($PAR_{max}$-$PAR_{min}$)/MAXIMP)*itr
       C= ln($B_{min}$/$B_{max}$)/MAXIMP
       B= Bmax*exp(C*itr)
       RND= Random(0,1)
       If (RND<=PAR) //Pitch Adjusting
         x'(i)= x'(i) + Random(-B,B)
         x'(i)= min(max(x'(i),$x^L$),$x^U$)
       EndIf
  Else //random harmony
       x'(i) = Random($x^U$,$x^L$)
  EndIf
EndFor
Return x'
```

Algorithm 2: Pseudo code for improvising new harmony vector in IHS

$$BtW = \frac{SSE_{BestHarmony}}{SSE_{WorstHarmony}} \qquad (8)$$

The concept of Best-to-Worst was inspired from the fact that the words "best" and "worst" are part of the HS algorithm nomenclature. The algorithm basically tries to find the "best" solution among a set of solutions stored in HM by improvising new harmonies to replace those "worst" ones. At any time HM would contain a number of solutions including best solution and worst solution in terms of their stored quality measures, i.e. fitness function values. If the worst fitness value in HM is close to that of the best, then this basically indicate that the quality of all current harmony vectors are almost as good as that of the best. This is somewhat similar to GA-based training methods when the percentage of domination of a certain member in the population could be used to signal convergence. Such domination is measured by the existence of a certain fitness value among the population. The BtW value would range between zero and one where values close to one indicate that the average fitness of harmonies in the current HM is close to the current best; a measure of stagnation. From another perspective, the BtW ratio would actually indicate the size of the area of the search space that is currently being investigated by the algorithm. Thus values close to zero would indicate a large search area while values close to one would indicate smaller areas.

The modified version of the adapted IHS training algorithm is referred to as the HS-BtW training algorithm. The BtW ratio would be used for dynamically adjusting the values of PAR and B as well as determining the training termination condition. A threshold value BtW_{thr} controls the start of PAR and B dynamic change and as shown in Fig. 9. This is analogues to the dynamic setting for the parameters of IHS given earlier in Fig. 4. Setting BtW_{thr} to 1.0 would make the algorithm behave just like the classical HS such that PAR is fixed at PAR_{min} and B is fixed at B_{max}. The BtW_{thr} value is determined by calculating BtW of the initial HM vectors prior to training.

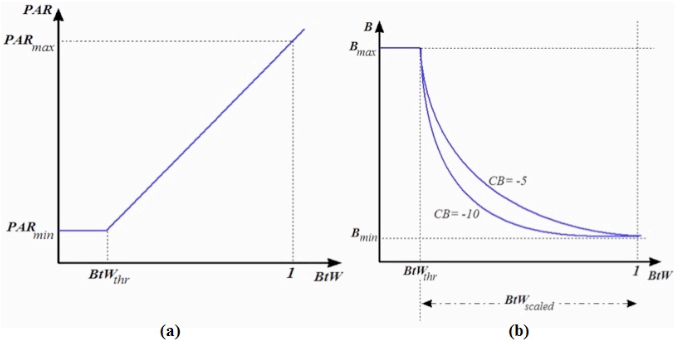


Figure 9. The dynamic PAR & B parameters of HS-BtW
(a) dynamic PAR value increases linearly as a function of the current HM BtW ratio, (b) dynamic B value decreases exponentially as a function of the current HM BtW ratio

PAR would be calculated as a function of the current BtW value and as given in equation (9) and (10) where *m* gives the line slop past the value of BtW_{thr}. B is also a function of the current BtW value and as given in equation (11) and (12) where *CB* is a constant controlling the steepness of change and it's in the range of [-10,-2] (based on empirical results) BtW_{scaled} is the value of BtW past the BtW_{thr} point scaled to be in the range [0,1].

The termination condition is based on $BtW_{termination}$ value that is set close to, but less than, unity. Training will terminate if BtW>= $BtW_{termination}$. MAXIMP is added as an extra termination criterion to limit the total number of training iterations if intended.

$$PAR(BtW) = \begin{cases} PAR_{\min} & \text{if } BtW < BtW_{thr} \\ m(BtW - 1) + PAR_{\max} & \text{if } BtW \geq BtW_{thr} \end{cases} \tag{9}$$

$$m = \frac{PAR_{\max} - PAR_{\min}}{1 - BtW_{thr}} \tag{10}$$

$$B(BtW) = \begin{cases} B_{\max} & \text{if } BtW < BtW_{thr} \\ (B_{\max} - B_{\min})\exp(CB \cdot BtW_{scalled}) + B_{\min} & \text{if } BtW \geq BtW_{thr} \end{cases} \tag{11}$$

$$BtW_{scalled} = \frac{(BtW - BtW_{thr})}{1 - BtW_{thr}} \tag{12}$$

where

BtW	Best-to-Worst ratio
BtW_{thr}	threshold value to start dynamic change
PAR_{min}	minimum pitch adjusting rate
PAR_{max}	maximum pitch adjusting rate
B_{min}	minimum bandwidth
B_{max}	maximum bandwidth
CB	constant controlling the steepness of B change

The flowchart shown in Fig. 10 shows the proposed HS-BtW training method, along with the pseudo code for improvising a new harmony vector in Algorithm 3. Both of these are analogous to adapted IHS flowchart given in Fig. 8 and improvisation process given in Algorithm 2. The IHS-based training method introduced earlier used two quality measures namely SSE and CEP where it was also indicated that SSE could be used as the sole fitness function. The HS-BtW method uses SSE only as its main fitness function in addition to using the BtW value as a new quality measure. Based on the BtW concept, the HS-BtW algorithm must compute this ratio in two places: after HM initialization process and after accepting a new harmony. The BtW value computed after HM initialization is referred to as BtW threshold (BtW_{thr}) used by equation (9) through (12). BtW is recomputed upon accepting a new harmony vector and the value would be used to dynamically set the value of PAR and B as well as to determine the termination condition. The HS-BtW improvisation process given in Algorithm 3 applies the newly introduced formulas given in equation (9) through (12).

F. *Initial parameter values*

The original IHS was used as an optimization method in many problems where the HMS value of 10 was encountered in many parameter estimation problems [61,9]. However it was indicated that no single choice of HMS is superior to others [16] and it is clear that in the case of FFANNs training more calculations would be involved if HMS were made larger.

HMCR was set to 0.9 or higher in many applications [58,59,57]. Based on the recommendations outlined by Omran et al [16], HMCR should be set such that HMCR≥0.9 for high dimensionality problems, which in this case resembles the total number of FFANN weights. It was also recommended to use relatively small values for PAR such that PAR≤0.5. The bandwidth B parameters values were selected based on several experimental tests in conjunction with selected $[x^L, x^U]$ range. Finally the termination condition would be achieved either if the value of $BtW \geq BtW_{termination}$, where $BtW_{termination}$ is set close to unity, or reaching the maximum iteration count specified by MAXIMP. Values like 5000, 10000 or higher were commonly used for MAXIMP in many applications [58,59,16].

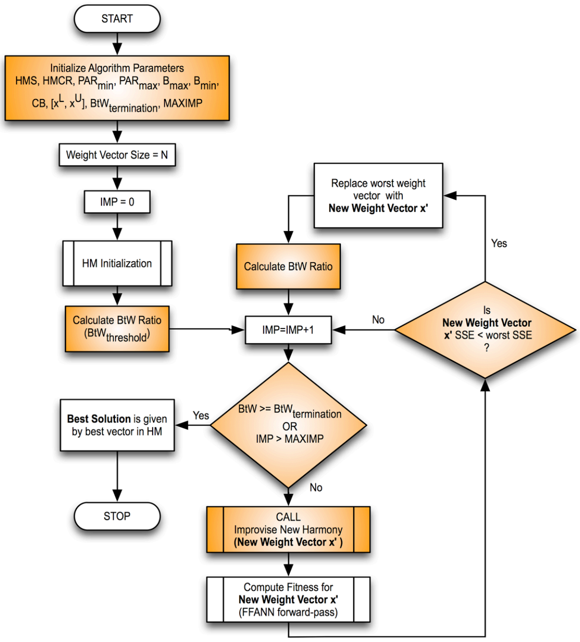


Figure 10. FFANN training using the HS-BtW algorithm

```
Create new harmony vector x' of size N
For i=0 to N do
  RND= Random(0,1)
  If (RND<=HMCR) //harmony memory considering
    RND= Integer(Random(0,HMS))
    x'(i)= HM(RND,i) //harmony memory access
    If (BtW<$BtW_{threshold}$)
      PAR= $PAR_{min}$
      B= $B_{max}$
    Else
      m= ($PAR_{max}$-$PAR_{min}$)/(1-$BtW_{threshold}$)
      PAR= m(BtW-1)+ $PAR_{max}$
      $BtW_{scaled}$= CB(BtW-$BtW_{threshold}$)/(1-$BtW_{threshold}$)
      B= ($B_{max}$- $B_{min}$)exp($BtW_{scaled}$)+ $B_{min}$
    EndIf
    RND= Random(0,1)
    If (RND<=PAR) //Pitch Adjusting
      x'(i)= x'(i) + Random(-B,B)
      x'(i)= min(max(x'(i),$x^L$),$x^U$)
    EndIf
  Else //random harmony
    x'(i) = Random($x^U$,$x^L$)
  EndIf
EndFor
Return x'
```

Algorithm 3: Pseudo code for improvising new harmony vector in HS-BtW

VI. Results and Discussion

In order to demonstrate the performance of the proposed methods, five different pattern-classification benchmarking problems were obtained from UCI Machine Learning Repository[1] [62] for the experimental testing and evaluation. The selected classification problems listed in Table (1) are taken from different fields including medical research, biology, engineering and astronomy. One of the main reasons behind choosing these data sets among many others is that they had no or very few missing input feature values. In addition these problems have been commonly used and cited in the neural networks, classification and machine learning literature [63-71]. All the patterns of a data set were used except for the Magic problem where only 50% out of the original 19,020 patterns of the data were used in order to perform the sequential computation within feasible amount of time. Some other pre-processing tasks were also necessary. For instance, in the Ionosphere data set there were 16 missing values for input attribute 6. These were encoded as 3.5 based on the average value of this attribute.

A 3-layer FFANN, represented by input-hidden-output units in Table 1, was designed for each to work as a pattern-classifier using the winner-take-all fashion [43]. The data set of each problem was split into two separate files such that 80% of the patterns are used as training patterns and the rest as out-of-sample testing patterns. The training and testing files were made to have the same class distribution, i.e. equal percentages of each pattern type. Data values of the pattern files where normalized to be in the range [-1,1] in order to suit the bipolar sigmoid neuron transfer function given in equation (6).

Table 1. Benchmarking data sets

Data Set	**Training Patterns**	**FFANN Structure**	**Weights**
Iris	150	4-5-3	43
Magic	9,510	10-4-2	54
Diabetes	768	8-7-2	79
Cancer	699	9-8-2	98
Ionosphere	351	33-4-2	146

For implementation Java 6 was used and all tests were run individually on the same computer in order to have comparable results in terms of the overall training time. The programs generate iteration log files to store each method's relevant parameters upon accepting an improvisation. The initial parameters values for each training method considered in this work are given in Table 2. GANNT and BP training algorithms were used for the training of the five aforementioned pattern-benchmarking classification problems to serve as a comparison measure against the proposed method.

The results for each of the benchmarking problems considered are aggregated in one table and are listed in Table 3 through Table 7. For each problem, ten individual training tests were carried out for each training method (M) considered. The best result out of the ten achieved by each method is reported for that problem. The aim is to train the network to obtain maximum overall recognition accuracy within the least amount of training time. Thus all comparisons consider the overall recognition accuracy as the first priority and the overall all training time as a second. The "Overall Time" in these tables represents the overall computing time required by each method to complete the whole training process. Some fields are not applicable for some methods and these are marked with (N.A.). For BP and GANNT the "Total Accepted" column represents the total number of training iterations needed by these methods.

Table 2. Initial parameter values used by training algorithms

M	Parameter	Values
IHS	HMS	10, 20
	HMCR	0.97
	PAR_{min}, PAR_{max}	0.1, 0.45
	B_{max}, B_{min}	5.0, 2.0
	$[x^L, x^U]$	[-250, 250]
	MAXIMP	5000, 20000
HS-BtW	HMS	10, 20
	HMCR	0.97
	PAR_{min}, PAR_{max}	0.1, 0.45
	B_{max}, B_{min}	5.0, 2.0
	CB	-3
	$[x^L, x^U]$	[-250, 250]
	$BtW_{termination}$	0.99
	MAXIMP	20000
GANNT	Population Size	10
	Crossover	At k=rand(0,N), if k=0 no crossover
	Mutation Probability	0.01
	Value Range [min,max]	[-250, 250]
	Stopping Criterion	50% domination of certain fitness
BP	Learning Rate	0.008
	Momentum	0.7
	Initial Weights	[-0.5, 0.5]
	Initialization Method	Nguyen-Widrow
	Stopping Criterion	SSE difference<= 1.0E-4

A. *The adapted IHS training method*

Since MAXIMP would determine the algorithm's termination condition, two values were used for testing, a lower value of 5000 and a higher value of 20,000. More iterations would give better chances for the algorithm to improvise more accepted improvisations. Results indicated by the IHS rows of Table 3 through 7 show that there are generally two trends in terms of the overall recognition percentage. In some problems, namely Magic, Diabetes and Ionosphere given in Table 4, 5 and 7 respectively, increasing MAXIMP would result in attaining better overall recognition percentage. The rest of the problems, namely Iris and Cancer given in Table 3 and 6 respectively, the resultant overall recognition percentage has decreased. Such case is referred to as "overtraining" or "overfitting" [43,72]. Training the network more than necessary would cause it to eventually lose its generalization ability to recognize out-of-sample patterns since it becomes more accustomed to the training set used. In general the best results achieved by the adapted IHS method are on par with those achieved by BP and GANNT rival methods. The IHS method scored best in the Iris, Cancer and Ionosphere problems given in Table 3, 6 and 7 respectively. BP scored best in the Magic problem given in Table 4 and GANNT scored best in the Diabetes problem given in Table 5.

[1] For full citations and data sets download see http://archive.ics.uci.edu/ml

Tests were also conducted using a double HMS value of 20. However, the attained results for all problems were the same as those attained using an HMS value of 10 but with longer overall training time and hence not reported in the results tables. For the problems considered in this work such result seem to coincide with that mentioned in [16] stating that no single choice of HMS is superior to others. Unlike the GA-based optimization methods, the HMS used by HS is different from that of the population size used in GANNT method. The HS algorithm and its dialects replace only the worst vector of HM upon finding a better one. Increasing the HMS would allow more vectors to be inspected but has no effect on setting the probabilistic values of both PAR and B responsible for the stochastic improvisation process and fine-tuning the solution. These values are directly affected by the current improvisation count and the MAXIMP value.

B. *The HS-BtW training method*

The adapted IHS method introduced in the previous section has achieved on par results in comparison with BP and GANNT. However, termination as well as the dynamic settings of PAR and B depended solely on the iteration count bounded by MAXIMP. The HS-BtW method has been used for the training of the same set of benchmarking problems using the same HMS value of 10. The results are given in the HS-BtW rows of Table 3 through 7. In comparison with IHS, BP and GANNT, the HS-BtW method scored best in the Iris, Diabetes and Cancer problems given in Table 3, 5 and 6 respectively. Sub-optimal results were obtained in the Magic and Ionosphere problems given in Table 4 and 7 respectively. However, due to its new termination condition and PAR and B settings technique, HS-BtW achieved convergence in much less number of total iterations and hence overall training time. The overall training time is the same as the last accepted improvisation time since termination occurs upon accepting an improvisation that yields BtW value equal or larger than $BtW_{termination}$.

Unlike the former adapted IHS, the HMS would have a direct effect on the HS-BtW performance since it affects the computed BtW ratio. Having a higher HMS would increase the solution space and the distance between the best solution and the worst solution. Tests were repeated using a double HMS value of 20 for all problems. The method attained the same results but with longer overall training time for Iris, Diabetes and Cancer problems given in Table 3, 5 and 6 respectively, and hence these were not included in the relevant results tables. This indicates that the HMS value of 10 is sufficient for these problems. However, HS-BtW was able to score higher in both the Magic problem and the Ionosphere problem given in Table 4 and 7 respectively when using an HMS value of 20. For the Magic problem, BP still holds the best score. The justifications for this is that BP has an advantage over the other considered methods when the training data set is relatively larger (see Table 1). Such increase in the number of training patterns will enable BP to have better fine-tuning attributed to its trajectory-driven approach. Table 8 summarizes the best results achieved by the IHS training method against those of the HS-BtW training method for the problems considered. For all the pattern-classification problems considered, the HS-BtW training method outperforms IHS in terms of the overall recognition percent and the overall training time even if double HMS is used by HS-BtW.

The convergence graph given in Fig. 11, which is obtained from the Iris results, illustrates how the BtW value changes during the course of training. Each drop in the “Worst Fitness” curve represent accepting a new improvisation that replaces the current worst vector of HM while each drop in the “Best Fitness” curve represent finding a new best vector in HM. The SSE value would decrease eventually and the two curves become close to each other as convergence is approached, i.e. as BtW value approaches $BtW_{termination}$. Fig. 12 shows the effect of BtW ratio on PAR and B dynamic settings. The top graph is a replica of lower graph of Fig. 11 which is needed to show the effect of BtW on PAR and B. The lower graph is a two-vertical axis graph to simultaneously show PAR and B changes against the upper BtW ratio graph. PAR would increase or decrease linearly with BtW as introduced earlier in Fig. 9(a). B on the other hand is inversely proportional with BtW and would decrease or increase exponentially as given earlier in Fig. 9(b). Such settings enables the method to modify its probabilistic parameters based on the quality of solutions in HM. In comparison with the adapted IHS method, the changes are steady and bound to the current iteration count to determine PAR and B values. In HS-BtW whenever the BtW value increases PAR values tend to become closer to the PAR_{max} value and B becomes closer to the B_{min} value. In the adapted IHS such conditions occurs only as the current iteration count approaches MAXIMP. The values of PAR and B would approach PAR_{min} and B_{max} respectively as the BtW values decreases. The horizontal flat curve area in the lower graph of Fig. 14 correspond to the case when the BtW values goes below the initial $BtW_{threshold}$. In this case, PAR is set fixed at PAR_{min} as in equation (9), while B is set fixed at B_{max} as in equation (11). Theses dynamic settings of the probabilistic parameters of PAR and B would gave the method better capabilities over the adapted IHS in terms of improvising more accepted improvisations in less amount of for the benchmarking problems considered.

VII. Conclusions

By adapting and modifying an improved version of HS, namely IHS, two new FFANN supervised training methods are proposed for pattern-classification applications; the adapted IHS and modified adapted IHS referred to as HS-BtW. The proposed IHS-based training methods has showed superiority in comparison with both a GA-based method and a trajectory-driven method using the same data sets of pattern-classification benchmarking problems. The settings of the probabilistic values in the adapted IHS training method are functions of the current iteration count. The termination condition is bound by a subjective maximum iteration count value MAXIMP set prior to starting the training process. Choosing a high value might cause the method to suffer from overtraining in some problems while choosing a smaller value might prevent the algorithm from finding a better solution. Increasing HMS seems to have no effect on the adapted IHS solutions for the pattern-classification problems considered for this work.

The HS-BtW method utilizes the BtW ratio to determine its termination condition as well as to dynamically set the probabilistic parameter values during the course of training. Such settings are independent of the current iteration count and have resulted in generating more accepted improvisations in less amount of overall training time in comparison with the adapted IHS. Doubling the HMS have resulted in attaining better solutions for some of the pattern-classification problems considered with an overall training time that is still less in comparison with other rival methods. However, BP is still superior in terms of attaining better overall recognition percentage in pattern-classification problems having relatively larger training data sets. BP seems to benefit from such sets to better fine-tune the FFANN weight values attributed to its trajectory-driven approach.

For future work it would be also interesting to apply the proposed HS-BtW technique to optimization problems other than ANNs such as some standard engineering optimization problems used in [15] or solving some global numerical optimization problems used in [30].

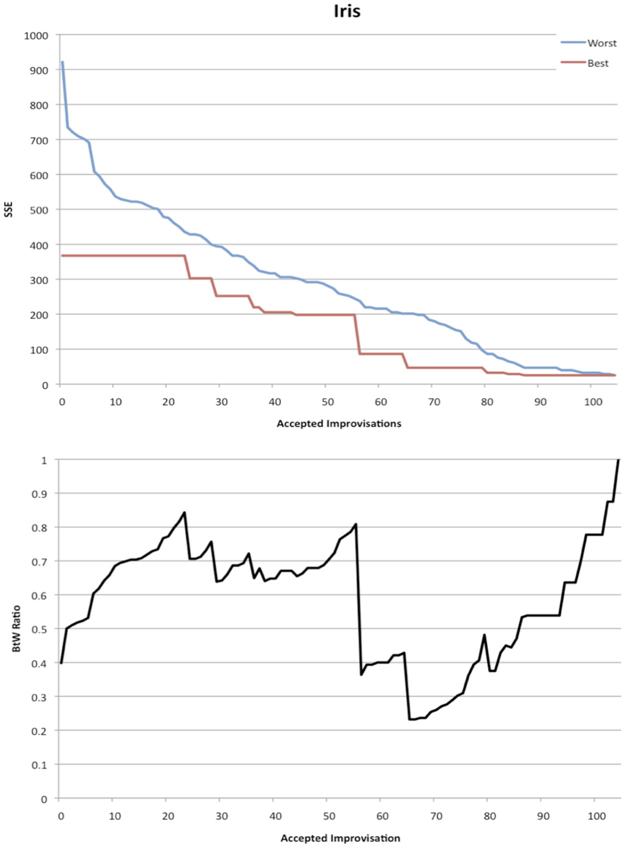


Figure 11. Convergence graph for the HS-BtW Iris problem

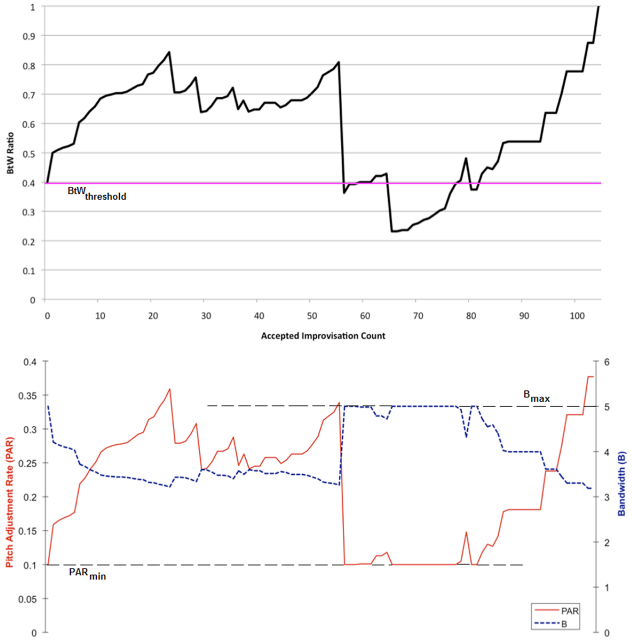


Figure 12. BtW value against PAR and B for the accepted improvisations of the HS-BtW Iris problem

ACKNOWLEDGMENT

The first author would like to thank Universiti Sains Malaysia for accepting as a postdoctoral fellow in the School of Computer Sciences. This work was funded by the Fundamental Research Grant Scheme from “Jabatan Pengajian Tinggi Kementerian Pengajian Tinggi” (Project Account number 203/PKOMP/6711136) awarded to the second author.

TABLE 3. RESULTS FOR BEST OUT OF TEN TRAINING SESSIONS FOR THE IRIS PROBLEM

M	**Training**							**Testing**	
	HMS/ Pop. Size	MAXIMP	SSE	Total Accepted	Last Accepted Iteration #	Last Accepted Time	Overall Time h:mm:ss	Overall Recog.%	Class Recog.%
IHS	10	5000	16	154	1826	0:00:58	0:02:39	96.67%	100.00% 100.00% 90.00%
	10	20000	7.08	287	10255	0:05:32	0:10:45	93.33%	100.00% 100.00% 80.00%
HS-BtW	10	20000	25.19	104	208	0:00:27	**0:00:27**	**100.00%**	100.00% 100.00% 100.00%
BP	N.A.	N.A.	7.85	1254	N.A.	N.A.	0:07:29	96.67%	100.00% 100.00% 90.00%
GANNT	10	N.A.	96	66	N.A.	N.A.	0:00:34	90.00%	100.00% 90.00% 80.00%

TABLE 4. RESULTS FOR BEST OUT OF TEN TRAINING SESSIONS FOR THE MAGIC PROBLEM

M	**Training**							**Testing**	
	HMS/ Pop. Size	MAXIMP	SSE	Total Accepted	Last Accepted Iteration #	Last Accepted Time	Overall Time h:mm:ss	Overall Recog.%	Class Recog.%
IHS	10	5000	12387.95	172	4574	1:49:43	1:59:13	77.39%	94.57% 45.74%
		20000	10647.98	413	19834	7:34:40	7:38:27	81.18%	93.27% 58.89%
HS-BtW	10	20000	11463.36	114	395	0:32:10	0:32:10	79.65%	86.62% 66.82%
	20	20000	9944.15	495	3190	4:10:01	4:10:01	81.44%	93.84% 58.59%
BP	N.A.	N.A.	6137.48	825	N.A.	N.A.	**4:35:42**	**83.97%**	82.97% 85.65%
GANNT	10	N.A.	12473.48	149	N.A.	N.A.	0:48:18	77.87%	89.62% 56.20%

TABLE 5. RESULTS FOR BEST OUT OF TEN TRAINING SESSIONS FOR THE DIABETES PROBLEM

M	**Training**							**Testing**	
	HMS/ Pop.Size	MAXIMP	SSE	Total Accepted	Last Accepted Iteration #	Last Accepted Time	Overall Time h:mm:ss	Overall Recog.%	Class Recog.%
IHS	10	5000	968	147	4835	0:10:48	0:11:10	76.62%	90.00% 51.85%
	10	20000	856	240	13001	0:27:11	0:41:47	77.27%	89.00% 55.56%
HS-BtW	10	20000	915.88	223	1316	0:11:42	**0:11:42**	**79.87%**	87.00% 66.67%
BP	N.A	N.A.	408.61	11776	N.A.	N.A.	5:30:42	78.57%	88.00% 61.11%
GANNT	10	N.A.	1108	1007	N.A.	N.A.	0:29:28	79.87%	89.00% 62.96%

TABLE 6. RESULTS FOR BEST OUT OF TEN TRAINING SESSIONS FOR THE CANCER PROBLEM

M	**Training**							**Testing**	
	HMS/ Pop.Size	MAXIMP	SSE	Total Accepted	Last Accepted Iteration #	Last Accepted Time	Overall Time h:mm:ss	Overall Recog.%	Class Recog.%
IHS	10	5000	124	155	4946	0:10:13	0:10:19	100.00%	100.00% 100.00%
	10	20000	99.76	212	19914	0:30:04	0:30:11	99.29%	100.00% 97.92%
HS-BtW	10	20000	126.37	217	1408	0:08:30	**0:08:30**	**100.00%**	100.00% 100.00%
BP	N.A.	N.A.	24.62	1077	N.A.	N.A.	0:27:55	95.71%	100.00% 87.50%
GANNT	10	N.A.	172	452	N.A.	N.A.	0:10:30	98.57%	100.00% 95.83%

TABLE 7. RESULTS FOR BEST OUT OF TEN TRAINING SESSIONS FOR THE IONOSPHERE PROBLEM

M	**Training**							**Testing**	
	HMS/ Pop.Size	MAXIMP	SSE	Total Accepted	Last Accepted Iteration #	Last Accepted Time	Overall Time h:mm:ss	Overall Recog.%	Class Recog.%
IHS	10	5000	72	181	4711	0:03:45	0:03:58	94.37%	100.00% 84.00%
	10	20000	64	225	19867	0:20:51	0:21:00	95.77%	97.83% 92.00%
HS-BtW	10	20000	113.6	327	1770	0:05:44	0:05:44	94.37%	100.00% 84.00%
	20	20000	70.23	584	7254	0:20:33	**0:20:33**	**97.18%**	100.00% 92.00%
BP	N.A.	N.A.	8.52	1628	N.A.	N.A.	0:24:43	95.77%	100.00% 88.00%
GANNT	10	N.A.	152	2244	N.A.	N.A.	0:35:57	94.37%	100.00% 84.00%

TABLE 8. IHS BEST TRAINING RESULTS VS. HS-BTW BEST TRAINING RESULTS

Problem	**IHS Training**			**HS-BtW Training**		
	HMS	Overall Time h:mm:ss	Overall Recog.%	HMS	Overall Time h:mm:ss	Overall Recog.%
Iris	10	0:02:39	96.67%	10	0:00:27	100.00%
Magic	10	7:38:27	81.18%	20	4:10:01	81.44%
Diabetes	10	0:41:47	77.27%	10	0:11:42	79.87%
Cancer	10	0:10:19	100.00%	10	0:08:30	100.00%
Ionosphere	10	0:21:00	95.77%	20	0:20:33	97.18%

REFERENCES

[1] Z. W. Geem, J. H. Kim, and G. V. Loganathan, "A New Heuristic Optimization Algorithm: Harmony Search", Simulation, vol. 72, pp. 60-68, 2001.

[2] Z. W. Geem, K. S. Lee, and Y. Park, "Applications of harmony search to vehicle routing", American Journal of Applied Sciences, vol. 2, pp. 1552-1557, 2005.

[3] Z. W. Geem, C.-L. Tseng, and Y. Park, "Harmony Search for Generalized Orienteering Problem: Best Touring in China," in Advances in Natural Computation. vol. 3612/2005: Springer Berlin / Heidelberg, 2005, pp. 741-750.

[4] Z. W. Geem, K. S. Lee, and C. L. Tseng, "Harmony search for structural design", in Genetic and Evolutionary Computation Conference (GECCO 2005), Washington DC, USA, 2005, pp. 651-652.

[5] R. Forsati, A. T. Haghighat, and M. Mahdavi, "Harmony search based algorithms for bandwidth-delay-constrained least-cost multicast routing", Computer Communications, vol. 31, pp. 2505-2519, 2008.

[6] R. Forsati, M. Mahdavi, M. Kangavari, and B. Safarkhani, "Web page clustering using Harmony Search optimization", in Canadian Conference on Electrical and Computer Engineering (CCECE 2008) Ontario, Canada: IEEE Canada, 2008, pp. 001601 – 001604.

[7] Z. W. Geem, "Harmony Search Applications in Industry," in Soft Computing Applications in Industry. vol. 226/2008: Springer Berlin / Heidelberg, 2008, pp. 117-134.

[8] W. S. Jang, H. I. Kang, and B. H. Lee, "Hybrid Simplex-Harmony search method for optimization problems", in IEEE Congress on Evolutionary Computation (CEC 2008) Trondheim, Norway: IEEE, 2008, pp. 4157-4164.

[9] H. Ceylan, H. Ceylan, S. Haldenbilen, and O. Baskan, "Transport energy modeling with meta-heuristic harmony search algorithm, an application to Turkey", Energy Policy, vol. 36, pp. 2527-2535, 2008.

[10] J.-H. Lee and Y.-S. Yoon, "Modified Harmony Search Algorithm and Neural Networks for Concrete Mix Proportion Design", Journal of Computing in Civil Engineering, vol. 23, pp. 57-61, 2009.

[11] P. Tangpattanakul and P. Artrit, "Minimum-time trajectory of robot manipulator using Harmony Search algorithm", in 6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON 2009) vol. 01 Pattaya, Thailand: IEEE, 2009, pp. 354-357.

[12] Z. W. Geem, "Novel Derivative of Harmony Search Algorithm for Discrete Design Variables", Applied Mathematics and Computation, vol. 199, pp. 223-230, 2008.

[13] A. Mukhopadhyay, A. Roy, S. Das, S. Das, and A. Abraham, "Population-variance and explorative power of Harmony Search: An analysis", in Third International Conference on Digital Information Management (ICDIM 2008) London, UK: IEEE, 2008, pp. 775-781.

[14] Q.-K. Pan, P. N. Suganthan, M. F. Tasgetiren, and J. J. Liang, "A self-adaptive global best harmony search algorithm for continuous optimization problems", Applied Mathematics and Computation, vol. 216, pp. 830-848, 2010.

[15] M. Mahdavi, M. Fesanghary, and E. Damangir, "An Improved Harmony Search Algorithm for Solving Optimization Problems", Applied Mathematics and Computation, vol. 188, pp. 1567-1579, 2007.

[16] M. G. H. Omran and M. Mahdavi, "Globel-Best Harmony Search", Applied Mathematics and Computation, vol. 198, pp. 643-656, 2008.

[17] D. Zou, L. Gao, S. Li, and J. Wu, "Solving 0–1 knapsack problem by a novel global harmony search algorithm ", Applied Soft Computing, vol. 11, pp. 1556-1564, 2011.

[18] R. S. Sexton and R. E. Dorsey, "Reliable classification using neural networks: a genetic algorithm and backpropagation comparison", Decision Support Systems, vol. 30, pp. 11-22, 15 December 2000.

[19] K. P. Ferentinos, "Biological engineering applications of feedforward neural networks designed and parameterized by genetic algorithms", Neural Networks, vol. 18, pp. 934-950, 2005.

[20] R. E. Dorsey, J. D. Johnson, and W. J. Mayer, "A Genetic Algoirthm for the Training of Feedforward Neural Networks", Advances in A.pngicial Intelligence in Economics, Finance, and Management vol. 1, pp. 93-111, 1994.

[21] J. Zhou, Z. Duan, Y. Li, J. Deng, and D. Yu, "PSO-based neural network optimization and its utilization in a boring machine", Journal of Materials Processing Technology, vol. 178, pp. 19-23, 2006.

[22] M. Geethanjali, S. M. R. Slochanal, and R. Bhavani, "PSO trained ANN-based differential protection scheme for power transformers", Neurocomputing, vol. 71, pp. 904-918, 2008.

[23] A. Rakitianskaia and A. P. Engelbrecht, "Training Neural Networks with PSO in Dynamic Environments", in IEEE Congress on Evolutionary Computation (CEC '09) Trondheim, Norway: IEEE, 2009, pp. 667-673.

[24] H. Shi and W. Li, "Artificial neural networks with ant colony optimization for assessing performance of residential buildings", in International Conference on Future BioMedical Information Engineering (FBIE 2009): IEEE, 2009, pp. 379-382.

[25] C. Blum and K. Socha, "Training feed-forward neural networks with ant colony optimization: an application to pattern classification", in Fifth International Conference on Hybrid Intelligent Systems (HIS '05) Rio de Janeiro, Brazil, 2005, p. 6.

[26] Z. W. Geem, C.-L. Tseng, J. Kim, and C. Bae, "Trenchless Water Pipe Condition Assessment Using Artificial Neural Network", in Pipelines 2007, Boston, Massachusetts, 2007, pp. 1-9.

[27] A. Kattan, R. Abdullah, and R. A. Salam, "Harmony Search Based Supervised Training of Artificial Neural Networks", in International Conference on Intelligent Systems, Modeling and Simulation (ISMS2010), Liverpool, England, 2010, pp. 105-110.

[28] S. Kulluk, L. Ozbakir, and A. Baykasoglu, "Self-adaptive global best harmony search algorithm for training neural networks", Procedia Computer Science, vol. 3, pp. 282-286, 2011.

[29] N. P. Padhy, Artificial Intelligence and Intelligent Systems, 1st ed. Delhi: Oxford University Press, 2005.

[30] J.-T. Tsai, J.-H. Chou, and T.-K. Liu, "Tuning the Strucutre and Parameters of a Neural Network by Using Hybrid Taguchi-Genetic Algorithm", IEEE Transactions on Neural Networks, vol. 17, January 2006.

[31] W. Gao, "Evolutionary Neural Network Based on New Ant Colony Algorithm", in International Symposium on Computational Intelligence and Design (ISCID '08). vol. 1 Wuhan, China, 2008, pp. 318 - 321.

[32] S. Kiranyaz, T. Ince, A. Yildirim, and M. Gabbouj, "Evolutionary artificial neural networks by multi-dimensional particle swarm optimization", Neural Networks, vol. 22, pp. 1448-1462, 2009.

[33] C. M. Bishop, Pattern Recognition and Feed-forward Networks: MIT Press, 1999.

[34] X. Jiang and A. H. K. S. Wah, "Constructing and training feed-forwardneural networks for pattern classifcation", Pattern Recognition, vol. 36, pp. 853-867, 2003.

[35] F. Marini, A. L. Magri, and R. Bucci, "Multilayer feed-forward artificial neural networks for class modeling", Chemometrics and intelligent laboratory systems, vol. 88, pp. 118-124, 2007.

[36] T. Kathirvalavakumar and P. Thangavel, "A Modified Backpropagation Training Algorithm for Feedforward Neural Networks", Neural Processing Letters, vol. 23, pp. 111-119, 2006.

[37] K. M. Lane and R. D. Neidinger, "Neural networks from idea to implementation", ACM Sigapl APL Quote Quad, vol. 25, pp. 27-37, 1995.

[38] E. Fiesler and J. Fulcher, "Neural network classification and formalization", Computer Standards & Interfaces, vol. 16, pp. 231-239, July 1994.

[39] L. Fausett, Fundamentals of Neural Networks Architectures, Algorithms, and Applications. New Jersey: Prentice Hall, 1994.

[40] I.-S. Oh and C. Y. Suen, "A class-modular feedforward neural network for handwriting recognition", Pattern Recognition, vol. 35, pp. 229-244, 2002.

[41] A. T. Chronopoulos and J. Sarangapani, "A distributed discrete-time neural network architecture for pattern allocation and control", in Proceedings of the International Parallel and Distributed Processing Symposium (IPDPS'02), Florida, USA, 2002, pp. 204-211.

[42] Z. W. Geem and W. E. Roper, "Energy demand estimation of South Korea using artificial neural networks", Energy Policy, vol. 37, pp. 4049-4054, 2009.

[43] M. H. Hassoun, Fundamentals of Artificial Neural Networks. Massachusetts: MIT Press, Cambridge, 1995.

[44] D. Kim, H. Kim, and D. Chung, "A Modified Genetic Algorithm for Fast Training Neural Networks," in Advances in Neural Networks - ISNN 2005. vol. 3496/2005: Springer Berlin / Heidelberg, 2005, pp. 660-665.

[45] M. b. Nasr and M. Chtourou, "A Hybrid Training Algorithm for Feedforward Neural Networks ", Neural Processing Letters, vol. 24, pp. 107-117, 2006.

[46] J. N. D. Gupta and R. S. Sexton, "Comparing backpropagation with a genetic algorithm for neural network training", Omega, The International Journal of Management Science, vol. 27, pp. 679-684, 1999.

[47] B. Guijarro-Berdinas, O. Fontenla-Romero, B. Perez-Sanchez, and A. Alonso-Betanzos, "A New Initialization Method for Neural Networks Using Sensitivity Analysis", in International Conference on Mathematical and Statistical Modeling Ciudad Real, Spain, 2006, pp. 1-9.

[48] J. Škutova, "Weights Initialization Methods for MLP Neural Networks", Transactions of the VŠB, vol. LIV, article No. 1636, pp. 147-152, 2008.

[49] G. Wei, "Study on Evolutionary Neural Network Based on Ant Colony Optimization", in International Conference on Computational Intelligence and Security Workshops Harbin, Heilongjiang, China, 2007, pp. 3-6.

[50] Y. Zhang and L. Wu, "Weights Optimization of Neural Networks via Improved BCO Approach", Progress In Electromagnetics Research, vol. 83, pp. 185-198, 2008.

[51] J. Yu, S. Wang, and L. Xi, "Evolving artificial neural networks using an improved PSO and DPSO", Neurocomputing, vol. 71, pp. 1054-1060, 2008.

[52] M. N. H. Siddique and M. O. Tokhi, "Training neural networks: backpropagation vs. genetic algorithms", in International Joint Conference on Neural Networks (IJCNN '01), Washington, DC 2001, pp. 2673 - 2678.

[53] K. E. Fish, J. D. Johnson, R. E. Dorsey, and J. G. Blodgett, "Using an Artificial Neural Network Trained with a Genetic Algorithm to Model Brand Share ", Journal of Business Research, vol. 57, pp. 79-85, January 2004 2004.

[54] E. Alba and J. F. Chicano, "Training Neural Networks with GA Hybrid Algorithms," in Genetic and Evolutionary Computation (GECCO 2004). vol. 3102/2004: Springer Berlin / Heidelberg, 2004, pp. 852-863.

[55] L. G. C. Hamey, "XOR Has No Local Minima: A Case Study in Neural Network Error Surface Analysis", Neural Networks, vol. 11, pp. 669-681, 1998.

[56] R. Cutchin, C. Douse, H. Fielder, M. Gent, A. Perlmutter, R. Riley, M. Ross, and T. Skinner, The Definitive Guitar Handbook, 1st ed.: Flame Tree Publishing, 2008.

[57] K. S. Lee and Z. W. Geem, "A New Meta-heuristic Algorithm for Continuous Engineering Optimization: Harmony Search Theory and Practice", Computer Methods in Applied Mechanics and Engineering, vol. 194, pp. 3902-3933, 2005.

[58] Z. W. Geem, "Optimal Cost Design of Water Distribution Networks Using Harmony Search", Engineering Optimization, vol. 38, pp. 259-277, 2006.

[59] Z. W. Geem and J.-Y. Choi, "Music Composition Using Harmony Search Algorithm," in Applications of Evolutionary Computing. vol. 4448/2007: Springer Berlin / Heidelberg, 2007, pp. 593-600.

[60] R. S. Sexton, R. E. Dorsey, and N. A. Sikander, "Simultaneous Optimization of Neural Network Function and Architecture Algorithm", Decision Support Systems, vol. 30, pp. 11-22, December 2004 2004.

AUTHORS PROFILE

Ali Kattan, (Ph.D.): Dr. Kattan is a postdoctoral fellow at the School of Computer Sciences - Universiti Sains Malaysia. He completed his Ph.D. from the same school in 2010. He has a blended experience in research and industry. Previously, he served as an assigned lecturer at the Hashemite University in Jordan and as a senior developer working for InterPro Global Partners, an e-Business solution provider in the United States. He specializes in Artificial Neural Networks and Parallel & Distributed Processing. His current research interests include optimization techniques, parallel processing using GPGPU, Cloud Computing and the development of smart phone application. Dr. Kattan is an IEEE member since 2009 and a peer-reviewer in a number of scientific journals in the field

Rosni Abdullah (Ph.D.): Prof. Dr. Rosni Abdullah is a professor in parallel computing and one of Malaysia's national pioneers in the said domain. She was appointed Dean of the School of Computer Sciences at Universiti Sains Malaysia (USM) in June 2004, after having served as its Deputy Dean (Research) since 1999. She is also the Head of the Parallel and Distributed Processing Research Group at the School since its inception in 1994. Her main interest lies in the data representation and the associated algorithms to organize, manage and analyse biological data which is ever increasing in size. Particular interest is in the development of parallel algorithms to analyse the biological data using Message Passing Interface (MPI) on message passing architectures and multithreading on multicore architectures. Her latest research interests include Cloud Computing, GPGPU and Computational Neuroscience.

QoS for Virtual Reality Software Based on RTCP over the Protocols of IP/UDP/RTP

ALBELAIHY ABDULLAH ABDULAZIZ, ALATEEBY SAAD MOHMAD, ABDUL NASIR BIN ZULKIFLI
Information Technology Department, UUM College of Arts and Sciences,
Universiti Utara Malaysia, 06010 Kedah, Sintok, MALAYSIA
abu_meteb30@hotmail.com, nasirzul@uum.edu.my

Abstract— **Current virtual reality environments are mainly visual experiences, displayed either on a computer screen or through special displays. Providing efficient and reliable network communication solution of virtual reality software has become an increase challenge for the industry experts and researchers, especially regards the aspects of the communication quality over the network. Thus, this paper intends to explore the possibility of using RTCP protocol with existing Panoweaver virtual reality software that uses IP/UDP/RTP communication protocol, in order to provide reliability, using RTCP could improve the existing software and Quality of Service (QoS) in terms of packet loss on increasing bandwidth. Therefore the amount of packet loss during the data transmission will be reduced respectively. Then, network simulation is used in this paper to compare the performance of the Panoweaver software with the Panoweaver-RTCP. This paper has the potentials to be much similar in the real time and is highly scalable and reliable for existing virtual reality software.**

***Keywords:** Virtual Reality, Panoweaver, RTP, IP, UDP, RTCP, QoS, NS2.*

I. INTRODUCTION

Nowadays, there has been a growth in concern in the prospective social impact of new technologies. In general, virtual world use widely for reflecting the self experience towards describing the variety of applications that commonly associated with immersive, highly visual, 3D [1]. Virtual world environments are designed based on applying a definite architecture for allowing network communication among parties, which usually divides the system into small components to process orders/requests from components [2]. This greatly helps to provide an advance services via the distributed processing mechanism. In addition, this distributed design can easily be adapted to work with any advanced network infrastructure for better transmission and delivery of data. An example of virtual reality software is the Panoweaver that use for creating the tour building by allowing stitching of 360 panoramas and publishing the panoramas to a single virtual reality tour showing the 360 degree view of the place. There are three main components that assist Panoweaver virtual reality software to be functioned and distributed among the network such as:

A. Server

The server component which is regarded as the main component of Panoweaver virtual reality software, acts as the controller. Using predefined rules upon conditions and statuses help to manage the communication among the network parties. However, it‘s invisible to users, as the system runs transparently. The system administrator is able to allow proper coordination of numerous sessions of queries and control between clients requests.

B. Client

The client components consist of individual PCs configured with suitable audio and video capturing or playback hardware. The client component comprises of six subcomponents:

i. Interface module (Virtual representation)
ii. Communication module
iii. Communication Control module
iv. Compression/Decompression module
v. Video module
vi. Audio module

C. Data Decompression

Data compression helps Panoweaver virtual reality software to keep continuous stream, which requires to be implemented on a separate machine.

Unfortunately, the current Panoweaver virtual reality software does not provide Quality of Service (QoS) for definite transmission and it relies on Real-time Transport Protocol (RTP) for delivering video and audio content through the Internet during the communication between the Panoweaver virtual reality users.

II. RTP CONTROL PROTOCOL (RTCP)

Some researchers as in [5] explained the RTP Control Protocol (RTCP) as a companion control protocol for RTP which used to provide out-of-band control information for an RTP flow as shown in the next figure and the reception quality feedback, participant identification, and the synchronization between media streams. RTCP runs alongside RTP and provides periodic reporting of this information [3]. Although data packets are typically sent every few milliseconds, the control protocol operates on the scale of seconds. The information sent in RTCP is necessary for synchronization between media stream; i.e., for lip synchronization between (audio and video). The RTCP's two main functions are:

i. It provides feedback on the quality of the media distribution. This function is performed by RTCP receiver and sender reports.

ii. For each sender, RTCP maps RTP time stamps for each RTP stream to a common sender clock [4], which allows audio and video synchronization on the receivers.

The RTCP packets contain direct information for quality-of-service monitoring. The RTCP consumes about 5% of the total bandwidth [5]. The Sender Reports (SR) and Receiver Reports (RR) exchange information on packet losses, delay and delay jitter [6]. This information may be used to implement a TCP like flow control mechanism upon UDP at the application level using adaptive encodings. A network management tool may monitor the network load based on the RTCP packets without receiving the actual data or detect the faulty parts of the network. The RTCP packets carry also a transport-level identifier (called a canonical name) for RTP source, which is used to keep track of each participant. Source description packets may also contain other textual information (user's name, email address) about the source. Albeit the source of the RTP packets is already identified by the SSRC identifier, an application may use multiple RTP streams, which can be easily associated with this textual information.

Novotny and Komosny (2007) introduced an optimization of hierarchical structure in SSM in order to achieve the lowest feedback transmission intervals, because the main issue in which the bandwidth is dedicated for the RTCP protocol. As defined in the RTP specification, it is limited to 5 % of the total allowed bandwidth and hence this creates a limiting factor for large-scale media streaming services based on Source-Specific Multicast-SSM (Figure 1) since the RTCP bandwidth is shared among all the receivers. As a result, noticeable larger delays 30 in sending feedback data from each receiver are encountered. However these noticeable larger delays have been curtailed by a hierarchical structure of receivers with summarization nodes.

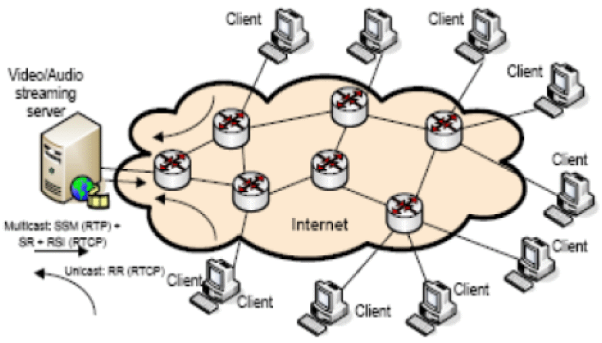


Figure 1: Example for SSM media streaming on internet [7]

While, Randa and Enugnla (2006) established their study according to different communication issues concerning the scalability of RTCP, such as; the delay of feedback and Bandwidth usage problem. They proposed a new scheme to tackle and reduce these problems. This scheme presents a hierarchical architecture, to organize the members dynamically in a hierarchy of local regions, and each region has (AG) aggregator, the members in every region will send their receiver reports RRs with limited scope to reach their aggregator AG. The AG aggregates statistics from the received receiver reports RRs and sends them to a Manager (AG-0) and the main function of the manager is to compute other statistics and evaluate the performance of network and eventually figuring out the regions with network congestion as shown in Figure 2. Even though, some issues were addressed such as congestion and overload due to the number of AGs for the manager (AG-0), and result is the burst of AG that are transmitted to the Manager (AG-0).

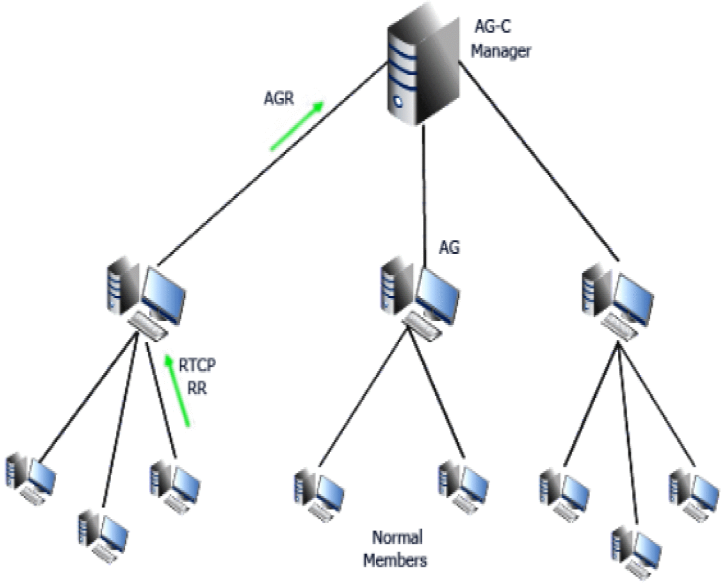


Figure 2: The scalable RTCP Architecture [8]

III. Simulation Environment

This paper test-bed was based on the Wide Area Network (WAN) with one Panoweaver virtual reality software (Demo) that links the clients with the server as shown in Figure 3. In our test scenario generated in the NS-2 simulators, Panoweaver server generates the client data and receives the processed data, presented in Figure 4.

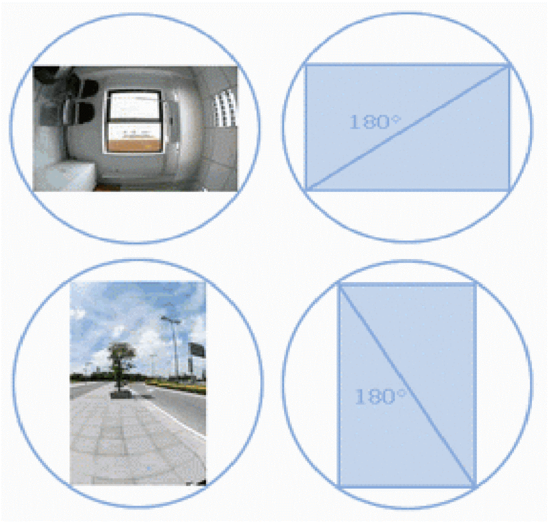


Figure 3: Panoweaver Dimensions

The definition of the sample time interval is the time elapsed between two consecutive receiver reports (RR). This can be expressed as the equation below:

$$\boxed{Network\_rate\_Loss = Network\_lost\ per\ Packet / Network\_recv * 100}$$

In RTCP for the Panoweaver software, RTCP was capable to adjust the frame rates of the generated packet that results in improvised the communication quality but at the same time results large bandwidth over the network. The graph below shows the following scenario for the network topology of Panoweaver.

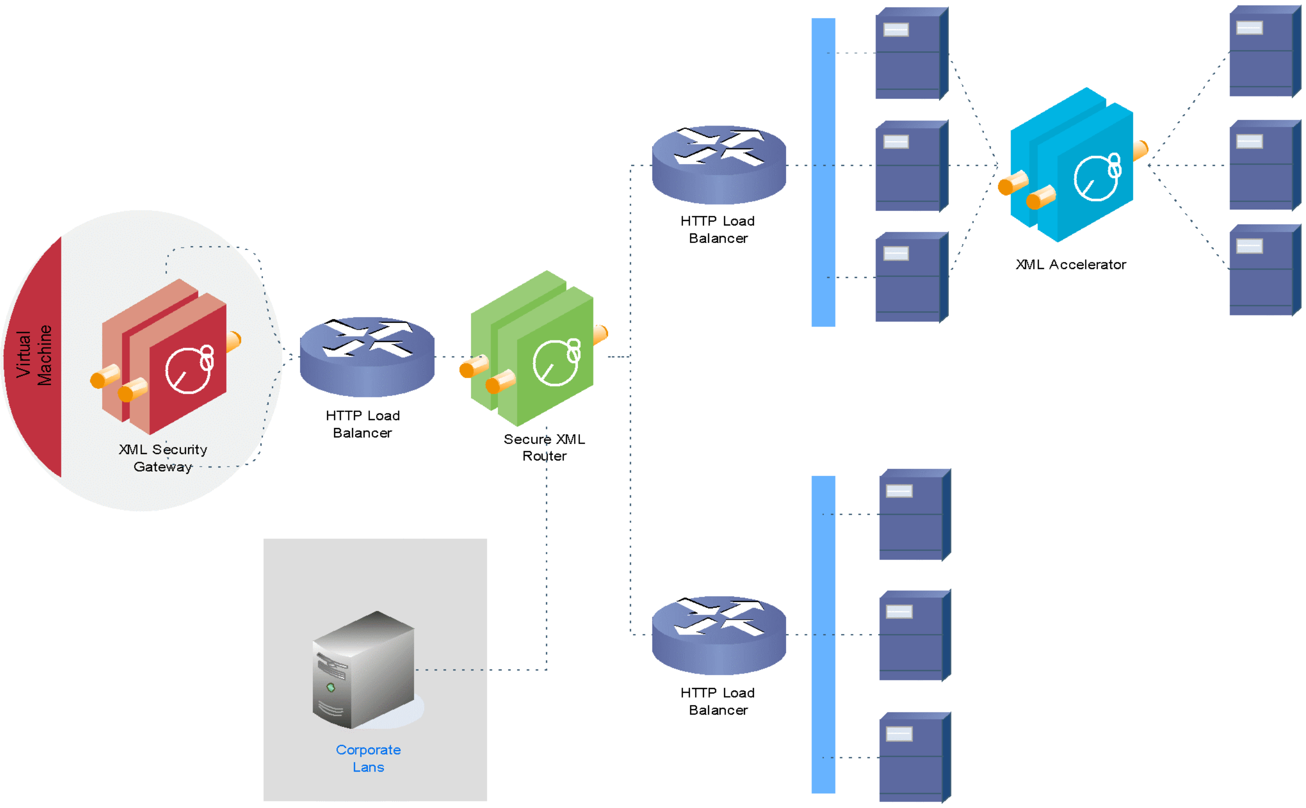


Figure 4: Simulated Network Topology of Panoweaver

IV. Evaluation: Network packet loss

Panoweaver evaluation was carried on one server that considered to be as a multicast to multiple numbers of clients. Normally in the packet transmission over network, the receiver has the capability to know the packet loss by calculating the difference or sequence of RTP sequence number that is present in the RTP communication message format. However, this paper simulated the Panoweaver by indicating the overall rates of the packet loss as a ratio of packet lost during the transmission over the packet received by the Panoweaver client. The Figure 5 illustrates the comparison of the packet loss in the Panoweaver and Panoweaver -RTCP communication protocol. But it is worth mentioning that although packet has been lost even with the utilization of RTCP feedback report, yet the

Panoweaver -RTCP will be able to adjust the frame rates of packet again whereas the normal Panoweaver do not.

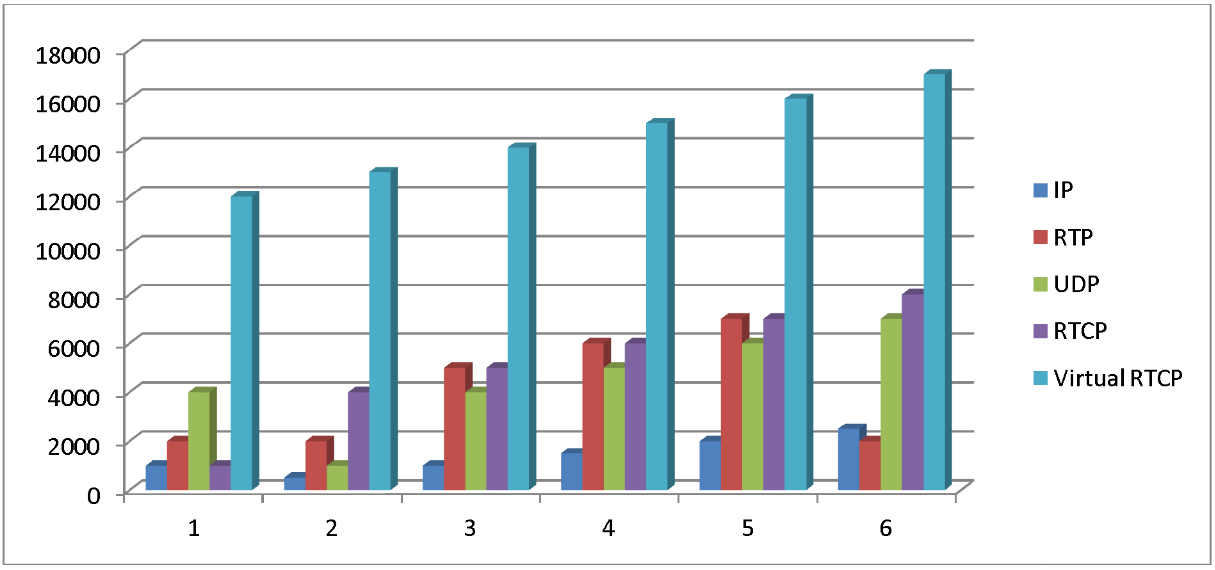


Figure 5: Packet Loss on Increasing Bandwidth

V. Conclusion

This paper was carried out to measure the performance of Panoweaver-RTCP virtual reality software for providing an efficient communication among user network over IP, UDP, and RTP. Decision feedback scheme, which is based on a RTCP, was designed by simulating the Panoweaver based on the packet loss rate. Implemented RTCP on top of UDP, where UDP assists the transmission of real time data, while RTCP provides feedback to sender and receiver about the transmission and reception of the media quality.

The approach of this paper has been justified by presenting a new architectural model on Panoweaver simulation using well suited network simulator NS-2. The finding indicated that Panoweaver gives better performance than the existing Panoweaver where the packet loss is expected.

VI. Acknowledgement

Acknowledgement for the Ministry of Higher Education in Kingdom of Saudi Arabia for the scholarship to carry out graduate studies, also to the Saudi Arabian Cultural Mission in Malaysia for support.

References

[1] G. Burdea and P. Coiffet, "Virtual reality technology," *Presence: Teleoperators & Virtual Environments*, vol. 12, pp. 663-664, 2003.

[2] T. Grantcharov, *et al.*, "Randomized clinical trial of virtual reality simulation for laparoscopic skills training," *British Journal of Surgery*, vol. 91, pp. 146-150, 2004.

[3] H. Gharavi, *et al.*, "MM06-8 RTCP-based Frame-Synchronized Feedback Control for IP-Video Communications over Multipath Fading Channels," in *International Conference on Communications*, USA 2004, pp. 1512-1516.

[4] Y. J. Liang, *et al.*, "Analysis of packet loss for compressed video: Effect of burst losses and correlation between error frames," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, pp. 861-874, 2008.

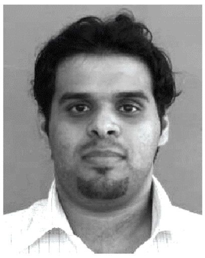
[5] J. Ott and J. Chesterfield, "E. Schooler," RTP Control Protocol (RTCP) Extensions for Single-Source Multicast Sessions with Unicast Feedback," RFC 5760, February2010.

[6] A. R. Reibman, *et al.*, "Quality monitoring of video over a packet network," *Multimedia, IEEE Transactions on*, vol. 6, pp. 327-334, 2004.

[7] V. Novotny and D. Komosny, "Optimization of large-scale RTCP feedback reporting in fixed and mobile networks," in *Third International Conference on Wireless and Mobile Communications*, UK, pp. 85 - 85., 2007, pp. 85-85.

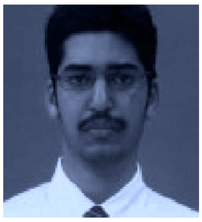
[8] E. I. M. Randa and M. Enugnla, "Enhanced Qos for Real-time Multimedia Delivery over the Wireless Link using RFID Technology," in *IEEE International Symposium* US, pp. 728 – 734., 2006.

LIST OF AUTHORS



Albelaihy Abdullah Abdulaziz

Is a master student at the Information and Communication Technology department, University Utara Malaysia. He is holding BSc (Computer), Teachers college, 2004/2005.



Alateeby saad mohmad

Is a master student at the Information and Communication Technology department, University Utara Malaysia. He is holding a BSc (Mathematic), King Abdulaziz University, 2008/2009.

Abdul Nasir Bin Zulkifli

Is working as Associate Professor at the College of Arts and Sciences. He is holding
MSc (Computer Integrated Manufacture), Loughborough University, 1993. BSc (Mechanical Engineering), Kansas State University, 1985.

ANEW APPROACH ON K-MEANS CLUSTERING

Trilochan Rout[1], Srikanta Kumar mohapatra [2], Jayashree Mohanty [3],
Sushant Ku. Kamillla [4], Susant K. mohapatra[5]

1,2,3- Computer Science and Engineerinmg Dept.,NMIET, Bhubaneswar,Oissa,India
4- Dept of Physics,ITER,Bhubaneswar,orissa,India
5- Chemical and Materials Engineering/MS 388, University of Nevada, Reno, NV 89557, USA

Abstract— **To explore the application of feature extraction technique to extract necessary features using k-mean clustering . The main goal of research on feature extraction using k-mean is to find out best features from the cluster analysis. All the implementation can be performed by using Genetic algorithm(GA) also. The same problem is done by using Mat lab. The k-mean clustering process for feature extraction gives accuracy almost equal with that Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).Although this is a unsupervised learning method, before classification of dataset into different class this method can be used to partition the group to obtain the better efficiency with respect to the number of object and attributes this can be developed with same logic and can give better accuracy in Genetic algorithm(GA).**

Keywords-: Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Genetic algorithm(GA).

I. INTRODUCTION

The need to understand large, complex, information-rich data sets is common to virtually all fields of business, science, and engineering. In the business world, corporate and customer data are becoming recognized as a strategic asset. The ability to extract useful knowledge hidden in these data and to act on that knowledge is becoming increasingly important in today's competitive world. So for the industries mining of data is important to take decision.

Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years, due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. The information and knowledge gained can be used for applications ranging from market analysis, fraud detection, and customer retention, to production control and science exploration. Mainly in statistical pattern classification this data mining is used. Statistical pattern classification deals with classifying objects into different categories, based on certain observations made on the objects. The possible information available about the object is in terms on certain measurements made on the object known as the features or the attribute set of the object.

In many applications, data, which is the subject of analysis and processing in data mining, is multidimensional, and presented by a number of features. The so-called "curse of dimensionality" pertinent to many learning algorithms, denotes the drastic raise of computational complexity and classification error with data having high amount of dimensions Hence, the dimensionality of the feature space is often reduced before classification is undertaken. Feature extraction and feature selection principles are used for reducing the dimension of the dataset. Feature extraction involves the production of a new set of features from the original features in the data, through the application of some mapping. Feature Selection involves the selection of important attributes or the features from the data set to make classify the data present in the data set.

Well-known unsupervised feature extraction methods include Principal Component Analysis (PCA) and k-mean clustering. The important corresponding supervised approach is Linear Discriminant Analysis (LDA).

Primary purpose of my work is to develop an efficient method of feature extraction for reducing the dimension. For this I have worked on new approach of k-mean clustering for feature extraction. This method extract the feature on the basis of cluster center.

1.2 Motivation

In the field of Data mining Feature Extraction has a tremendous application such as dimension reduction, pattern classification, data visualization, Automatic Exploratory Data Analysis. To extract proper feature from the rich data set is the major issue. For this many work has been done before to reduce dimension. Mainly PCA and LDA are used for this dimension reduction. Identification of important attributes or features is a major area of research from last several years. To give new solution to some long standing necessities of feature extraction and to work with a new approach of dimension reduction. PCA finds a set of the most representative projection vectors such that the projected samples retain the most information about original samples. LDA uses the class information and finds a set of vectors that maximize the between-class scatter while minimizing the within-class scatter. Cluster is another technique for making group for the different object present in the dataset. With the cluster center also it can be possible to find out the necessary feature from the data set. In my present work I use this new approach of extracting the feature.

2. An Overview of Data Mining and Knowledge Discovery

Data mining is an iterative process within which progress is defined by discovery, through either automatic or manual methods. Data mining is most useful in an exploratory analysis scenario in which there are no predetermined notions about what will constitute an "interesting" outcome. Data mining is the search for new, valuable, and nontrivial information in large volumes of data. It is a cooperative effort of humans and computers. Best results are achieved by balancing the knowledge of human experts in describing problems and goals with the search capabilities of computers.

The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. A cluster of data objects can be treated collectively as one group and so may be considered as a form of data compression. Although classification is an effective means for distinguishing groups or classes of objects, it requires the often costly collection and labeling of a large set of training tuples or patterns, which the classifier uses to model each group. It is often more desirable to proceed in the reverse direction: First partition the set of data into groups based on data similarity (e.g., using clustering), and then assign labels to the relatively small number of groups. Additional advantages of such a clustering-based process are that it is adaptable to changes and helps single out useful features that distinguish different groups.

As a branch of statistics, cluster analysis has been extensively studied for many years, focusing mainly on *distance-based cluster analysis*. Cluster analysis tools based on *k*-means, *k*-medoids, and several other methods have also been built into many statistical analysis software packages.

In machine learning, clustering is an example of unsupervised learning. Unlike classification, clustering and unsupervised learning do not rely on predefined classes and class-labeled training examples. For this reason, clustering is a form of learning by observation, rather than *learning by examples*. In data mining, efforts have focused on finding methods for efficient and effective cluster analysis in *large databases*. Active themes of research focus on the *scalability* of clustering methods, the effectiveness of methods for clustering complex shapes and types of data, high-dimensional clustering techniques, and methods for clustering mixed numerical and categorical data in large databases.

Two types of clustering algorithms are ***nonhierarchical*** and ***hierarchical***. In nonhierarchical clustering, such as the **k-*means*** algorithm, the relationship between clusters is undetermined. Hierarchical clustering repeatedly links pairs of clusters until every data object is included in the hierarchy. With both of these approaches, an important issue is how to determine the similarity between two objects, so that clusters can be formed from objects with a high similarity to each other. Commonly, ***distance functions***, such as the ***Manhattan*** and ***Euclidian*** distance functions, are used to determine similarity. A distance function yields a higher value for pairs of objects that are less similar to one another. Sometimes a ***similarity function*** is used instead, which yields higher values for pairs that are more similar.

Data clustering is a common technique for statistical data analysis, which is used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics. The computational task of classifying the data set into *k* clusters is often referred to as ***k*-clustering**.

Simply speaking k-means clustering is an algorithm to classify or to group your objects based on attributes or features into K number of group. K is positive integer number. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Thus the purpose of K-mean clustering is to classify the data.

4.3 K-means algorithm :

The basic step of k-means clustering is simple. In the beginning we determine number of cluster K and we assume the centroid or center of these clusters. We can take any random objects as the initial centroids or the first K objects in sequence can also serve as the initial centroids. Then the K means algorithm will do the three steps below until convergence

Iterate until *stable* (= no object move group):

1. Determine the centroid coordinate
2. Determine the distance of each object to the centroids
3. Group the object based on minimum distance

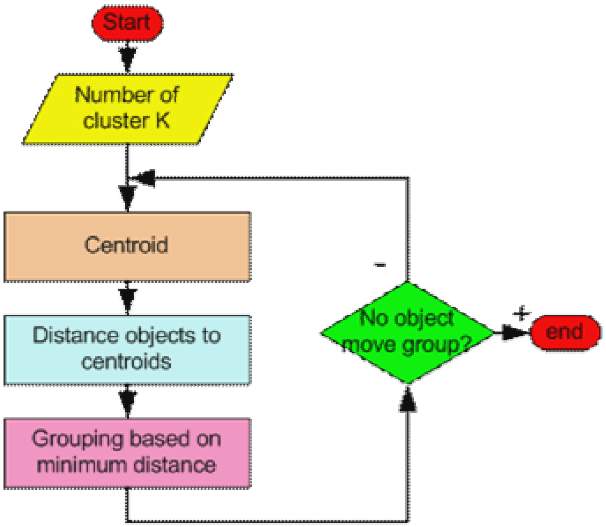

Start
Number of cluster K
Centroid
Distance objects to centroids
Grouping based on minimum distance
-
No object move group?
+
end

Fig 4.1: Flow chart for finding Cluster

1.**Initial value of centroids :** Assign the first k object as the initial cluster and their centroid can be found by assigining directly their attributes value initially.

2. **Objects-Centroids distance :** we calculate the distance between cluster centroid to each object with the help of Euclidean distance between points $P=(p_1,p_2,.....,p_n)$ and $Q=(q_1,q_2,.....,q_n)$ in Euclidean *n*-space, is defined as:

$$\sqrt{(p_1-q_1)^2+(p_2-q_2)^2+\cdots+(p_n-q_n)^2}=\sqrt{\sum_{i=1}^{n}(p_i-q_i)^2}.$$

3.Object Clustering : Assigning the object to that group or cluster if that object has having minimum distance with that cluster in compare to other cluster.

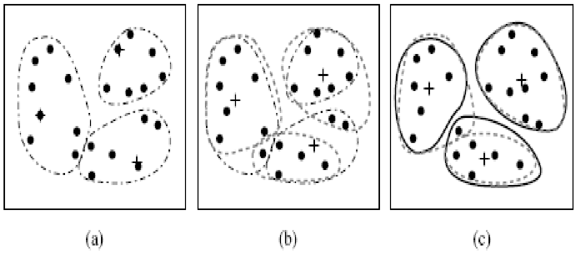


Fig 4.2 Clustering of a set of objects based on the *k*-means method. (The mean of each cluster is marked by a "+".)

The *k*-means method, however, can be applied only when the mean of a cluster is defined. This may not be the case in some applications, such as when data with categorical attributes are involved.

Future work:

The k-mean clustering process for feature extraction gives accuracy almost equal with that Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). With large number of record set the accuracy of k-mean is slightly degrades. Although this is a unsupervised learning method, before classification of dataset into different class this method can be used to partition the group To obtain the better efficiency with respect to the number of object and attributes this can be further developed with same logic in GA.

References:

1. Sushmita Mitra, Sankar K. Pal and Pabitra Mitra, "*Data Mining in Soft Computing Framework: A Survey*", in IEEE
2. Robert S. H. Istepanian, Leontios J. Hadjileontiadis, and Stavros M. Panas, "*ECG Data Compression Using Wavelets and Higher Order Statistics Methods*", IEEE Transactions on Information Technology In Biomedicine, Vol. 5, No. 2, June 2001.
3. T.W. Anderson. Asymptotic theory for principal component analysis. Annals of Mathematical Statistics, 34:122–148, 1963.
4. W.N. Anderson and T.D. Morley. Eigenvalues of the Laplacian of a graph. Linear and Multilinear Algebra,18:141–145, 1985.
5. W.E. Arnoldi. The principle of minimized iteration in the solution of the matrix eigenvalue problem. Quarterlyof Applied Mathematics, 9:17–25, 1951.
6. M. Balasubramanian and E.L. Schwartz. The Isomap algorithm and topological stability. Science, 295(5552):7,2002.
7. G. Baudat and F. Anouar. Generalized discriminant analysis using a kernel approach. Neural Computation,12(10):2385–2404, 2000.
8. M. Belkin and P. Niyogi. Laplacian Eigenmaps and spectral techniques for embedding and clustering. In Ad-vances in Neural Information Processing Systems, volume 14, pages 585–591, Cambridge, MA, USA, 2002. TheMIT Press.
9. A.J. Bell and T.J. Sejnowski. An information maximization approach to blind separation and blind deconvolution.Neural Computation, 7(6):1129–1159, 1995.
10. Y. Bengio, O. Delalleau, N. Le Roux, J.-F. Paiement P. Vincent, and M. Ouimet. Learning eigenfunctions linksspectral embedding and Kernel PCA. Neural Computation, 16(10):2197–2219, 2004.
11. Michail Vlachos, Jessica Lin, Eamonn Keogh and Dimitrios Gunopulos "*A Wavelet-Based Anytime Algorithm for KMeans Clustering of Time Series*", 3rd SIAM International Conference on Data Mining. San Francisco, CA. May 1-3, 2003, Workshop on Clustering High Dimensionality Data and Its Applications.

A Taxonomy of Malicious Programs For An End User

Muhammad Azhar Mushtaq
Departemnt of Computer Science and IT
University of Sargodha
Sargodha, Pakistan.
azhar.mushtaq@uos.edu.pk

Madiah Sarwar
Department of Computer science and IT
University of Sargodha
Sargodha, Pakistan
madiha.sarwar@uos.edu.pk

Abstract- **Computer and network attacks have become highly sophisticated and complex with different names and multiple characteristics. In order to understand and find solutions against new and old attacks, different types of computer and network taxonomies are utilized. However, such taxonomies are being actively developed for expert users; research efforts towards making attack taxonomy for basic end users are still isolated. In this work we present taxonomy for the end users that will help in identifying attacks, the precaution measures they need to adapt and how to categorize new attacks. Moreover, through an empirical survey of the taxonomy, it is concluded that end users will be more protected than before and validity of the taxonomy was also checked.**

Keywords-Computer and netwrok attack; taxonomy; end users

I. INTRODUCTION

Attacks on computers and networks have a long lasting history, which requires constant attention. Different attack techniques are carried out by attackers to fulfill their objectives. In the recent years they have spread more rapidly and since 1999 there is a marked increase in the number of incidents reported by Computer emergency response team (CERT). Moreover, in year 2008 F-secure managed to collect more than ten million suspicion samples [6] [7]. This situation is alarming and deep rooted and end user feel to be more insecure than any one else. One of the strongest reasons is that, in the beginning launching these attacks required relatively more technical knowledge and expertise but today they have become user friendly and their propagation is much faster and easier than ever before. It is therefore the need of the time to make aware not only the corporate or big business but end users working for these business and those sitting in homes to be well informative regarding these malicious attacks.

In order to answer all these serious concerns many taxonomies were proposed by the researchers and their sole purpose was to present and provide a meaningful way of classifying these attacks. Unfortunately, all the earlier taxonomies employ a unique way of classifying attacks. Some classify attacks by their distinctive names like virus, worm and others classify attacks according to the weakness in the system. Because of different classification schemes and categorizing attacks differently, it is not possible for end users to understand these attacks and it creates confusion in taking proper precautionary measures. Due to this fact, a new taxonomy model is proposed in this area for the betterment of end users. The proposed taxonomy is based on four distinctive aspects damage, cost, propagation, and precaution.

Every attack has some damaging effects, some attacks may cause severe damages and some may have no damaging effect. For example, a virus may cause damage at computer level by infecting hardware or other parts of it but cannot damage the network; where as a simple worm with no extra threat only attacks the network by overloading it. Cost is the second aspect through which a user can classify or understand attacks. Cost can be referred to in two ways; cost of damages and cost of fixing these damages. Most attack types have some kind of propagation mechanism, i.e. they try to replicate themselves and spread. In many cases the propagation depends upon human interaction with them. In case of a virus, propagation will not take place until it comes in contact with an end user. On the other hand, a worm spreads by itself. Precaution is most important part of the taxonomy, because this can be used in classifying attacks and it will keep end users protected from attacks. Precaution must be taken on two levels; one is the administration level and second is the end user level. Administration level precautions are not discussed here in detail because administrators already have the knowledge and skills to protect the network. The end user must take certain precautions on their personal computer in order to keep the computer safe from attacks.

The remainder of this paper is organized as follows. Some of the previous related taxonomies are reviewed in section 2. Section 3 presents empirical survey of the taxonomy where as proposed taxonomy model is covered in section 4. Section 5 concludes the paper and present future work.

II. RELATED WORK

In the following section some of the prominent taxonomies are presented.

A. Taxonomy based on Computer Vulnerabilities

1) Protection analysis report 1978

In 1978, Information Science Institute at University of Southern California launched project called Protection Analysis (PA). It was an effort to sort errors in operating system, applications and discover techniques which can detect weaknesses in software errors [1]. The PA report first came up with ten categories but after further the numbers of categories were reduced to four global errors: domain errors, validation error, naming error, and serialization error.

2) Bishop taxonomy

In 1995, Bishop presented his vision of a taxonomy which was different from the previous taxonomies. His work includes vulnerabilities in UNIX and the classification schemes were based on the basics of these vulnerabilities. Bishop presented his taxonomy in the form of 6 axes (Nature, Time of introduction, Exploitation domain, Effect domain, Minimum number, Minimum number and Source) [2].

B. Taxonomy based on Computer Attacks

1) Landwehr et al., taxonomy

Landwehr presented their taxonomy on computer programs and security flaws along with 50 actual flaws. As earlier taxonomies collected data during the development of the software Landwehr paid attention to the security flaws that happen after the software is released for use. Landwehr taxonomy mainly emphasize on organizing flaws, adding new ones and users can get information on which part of the system is causing more trouble. The flaws were broken down on the basis of genesis (how), time of introduction (when), and location (where). These three categories are explained in detail in the next section [3].

a) Origin of flaw

The important part in this section is the method through which security flaw is inserted into the system. First find out whether it was done by proper planning or it happened accidentally. Landwehr argued that sometimes this could be confusing because program like remote debugging have deliberately given functions which at the same time can provide unintentional security flaws.

The next category is the harmfulness of the flaws. Damaging flaws contain trojan horse, trapdoor, and logic bomb; these threats can further be classified in duplicating and non-duplicating threats. Another category under intentional flaw is covert channels which transfer information against the will of the system designer [3].

b) Time of introduction

To find exactly when the flaw was introduced during software development, Landwehr proposed the second stage called time of introduction which was further divided into three components: development, maintenance, and operation. During the development phase different implementations are done in order to meet certain conditions. If these implementations are not properly done there are chances of a flaw being activated. Programmers can make different mistakes in these activities such as not complying with the terms of software requirements during source coding. Maintenance is the time when the software is released but still being used on testing purposes. Landwehr pointed out that during the maintenance time programmers usually fix a flaw but do not track it back to the source, this could awake more flaws. Moreover, due to viruses or unauthorized access there could be changes done in the software during the operation time. Operation time is when the software is out in the market and organizations are using them [3].

c) Location

The third phase in the taxonomy was the location of the flaw. The location was divided in two parts, software and hardware. Because mainly emphasis was on software, so it was further divided into operating system, support software, and application software. Some of the flaws under operating system can take place if the system did not accurately initialized the defense measure or an outsider gain admittance because of a fault in memory management [3].

2) Howard Taxonomy

Howard presented in his PhD thesis the taxonomy of computer and network attacks. His taxonomy was based on the trail an attack goes along rather than the security flaws. His process-based taxonomy consists of five stages: attackers, tools, access, results and objectives [4].

An attacker could be any one who purposefully cracks into a computer. Attackers could be different types of people such as hackers, terrorists, and vandals. These attackers utilize some form of tools in order to get admittance. Variety of tools is available, ranging from user command to data tapping. By using the vulnerabilities in implementation, design, and configuration an attacker can get access. The results of this can be corruption of information, disclosure of information or denial of service. Through this process the attackers accomplish the objectives which can be financial or political gain. This process based taxonomy is very useful for understanding how the attack process works. However, if motivation and objectives are not given any importance this taxonomy is not valuable. Howard and Thomas (1998) made changes in the process-based taxonomy but failed in fulfilling the requirements [4].

3) Hansman Taxonomy

Hansman criticized on Howard's taxonomy because it explains the attack process and does not clarify attacks which happen on daily basis. For example the Code Red worm cannot be classified using the Howard taxonomy. Hansman's approach was to categorize computer attacks such as virus, worms, and trojans; attacks which a user faces every day. Also, Hansman wanted a taxonomy in which attacks with multiple threats (blended attacks) can be classified. For these reasons Hansman proposed a new taxonomy which consists of dimensions [5].

a) First dimension

In the first dimension attacks are classified by attack vectors. Attack vector is the way attackers gain access to their targets so that certain payloads or harmful contents can be transported. It provides the path for hackers to break into a system or network; it can also give exact information about an attack. For example, Melissa virus propagates through e-mail so according to first dimension it is considered as mass-mailing worm [5] [8].

b) Second dimension

Second dimension is based on the attack targets. If attack has more than one target, more than one entry can be made in this dimension. For example, if Server A is attacked targets would be operating system and service rather then the server. In case Code Red attacks server A, the target would be Internet Information Server (IIS) and not Server A itself [5].

c) Third dimension

Third dimension is based on the vulnerabilities that an attack exploits. If attack utilizes more then one vulnerability, there could be multiple entries in third dimension. As Common Vulnerabilities and Exposures (CVE) provides an easier and a general name for a weakness, that is why Hansman included it in his taxonomy. The CVE data sources strongly indicate the fact that Code Red worm can take advantage of the weakness in Microsoft internet information services. Hansman also proposed that in case the vulnerabilities are not found under CVE database then one of Howard's vulnerabilities should be selected. Howard three vulnerabilities were vulnerability in implementations, vulnerability in design, and vulnerability in configuration [5].

d) Fourth dimension

Hansman fourth dimension depends upon the payloads or effects which have extra features. Such as a worm may simply demolish some files and also have a trojan payload at the same time. Hansman further discussed that the taxonomy can be improved by adding more dimensions [5].

III. EMPERICAL SURVEY

Before proposing the taxonomy, a survey was conducted in order to measure the awareness level about computer attacks and the threat level among end users in Pakistan .The sample of the study was taken from different university students from all over Pakistan. A total of 500 questioners were distributed randomly among different universities students in Pakistan. Out of the 500 distributed 450 were useable for conducting further analysis.

The data sample was analyzed using SPSS statistical package and this can be a key element when proposing the taxonomy. The survey was divided in two sections. The first section covers demographic questions such as gender, age, qualification and etc. The demographic section is not included in this paper because for proposing taxonomy these demographic questions are irrelevant. The aim is to provide a computer attack taxonomy which can be beneficial for all end-users. The second section consists of statement questions which focus on the respondent's awareness, effect of computer attack and the precautions against such attacks. The survey questionnaire was designed based upon likert scale of 1-5 with 1 strongly disagreed to 5 strongly agreed. This method was used so that respondent's answers can be clear and no ambiguity between answers should rise.

The item reliability was measured using cornbach alpha which is type of internal reliability estimation used to measure the consistency of responses on a composite measure that contains more than 1 item. The value closer to 1 is considered as a good measure. In our case the cornbach alpha values above .60 is considered acceptable. In the survey analysis values ranged between .65 to .78. The results of one sample t-test show high significance level <.001 on all the attributes. The overall mean value of attribute 1 damage is 2.64, which states that there exists a partial awareness of damage among the respondents. Similar results have been found on cost and propagation attributes having an overall mean value of 2.49 and 2.86. This indicates an alarming situation that end users have partial awareness about the cost and they have to pay in the shape of loss of losing there important data, confidential information, personal identity, etc. As far as precautionary measures are concerned against all kind of threats it has been seen that the level of awareness is moderate with the mean values ranging between 3.0 to 3.3 on all the attributes namely precaution against virus, worm, Trojan, spam and phishing. An inference that could be drawn is that the end users at one end have either zero or partial awareness about the consequences of threats while on the other end they have prepared themselves against these threats at quite a moderate precautionary level. According to tabel 1 the conclusion can be drawn depending on the mean value of each question about whether the end user posses high awarness (H.A), moderate awarness (M.A) or partial awarness (P.A) about each questionaaire. It is worth mentioning here that end users are not aware of what kind of protection they might need against different type of threats.

IV. TAXONOMY MODEL

The attacks are categorized according to their harmful purpose. The harmful purpose can be for example, damaging computer or network resources, stealing of confidential files, financial fraud, identity theft, etc. virus, worm, trojan horse, spam and phishing are the subcategories of a malware attack. Spam and phishing are both a part of spoofing which means lying about ones own identity. As these attacks have malicious purpose they are included in the category of malware attacks in the proposed taxonomy. In table 2 the taxonomy is explained in detail for end user benefit.

A. First aspect

Virus can damage both computers and networks. At computer level, the hardware damages are done to processor, hard disk, CD ROM and in software it can damage parts of application, file or the whole operating system. Virus cannot damage the network but utilizes the network in order to propagate [9]. Worms are different in means of damaging as they can install backdoors in the system that can then be remotely accessed by attackers. Worm usually uses up the whole network bandwidth for replicating purpose making the network to crash or slow down. With the help of trojans a attacker can view someone else's desktop, or can notice the input given to the system through key strokes loggers. It can also make changes in the BIOS (Basic input/output system) of the system, changing system settings and can even upload some kind of other malicious program such as virus or worm. Modification of the data is also the damaging effect of trojans [10]. Due to phishing users can lose all their financial information, credit card numbers, social security number, and bank account details. Phishing damages are mostly related to money because the motive of the attacker is to obtain financial information. Attackers use spam in order to freeze the network or computer by sending hundred to thousands of copies to each end user. It even consume up server disk space so even the legitimate e-mails cannot be delivered. This can cost money to companies' or organizations that heavily rely on business through e-mails.

B. Second Aspect

Cost of fixing the damages depend on what type of attack took place. In case of virus it can damage computer hardware as well as software and fixing these things cost money. But there are some other costs such as losing of important files which the end user has to retrieve, lost passwords, pictures, etc. In worms, by shutting down the network the worm will stop propagating. Shutting down the network has affects such as; money loss in business. Sometimes removing the worm can take weeks and the cost could go in millions of dollars. In trojans cost varies because trojan may install other malicious programs. In case of a simple trojan costs are as follow: money lost because of no service, confidential information stolen, time and money spent to restore computer settings back to normal condition. Phishing damaging costs are divided in two parts: cost to service providers and cost to end users. The service providers have to bear the cost of providing service to phishing victims, who call the companies to resolve fraud matters. In some cases companies have to block customer accounts, which is not good for business and the trust between customers and companies may no longer survive. As far as end users are concerned, the main cost is losing one's personal information. Personal information means bank detail, credit card information, and social security number. Other costs are tracking down the culprit behind the scheme, calling or meeting with different organizations to resolve the matter, reporting to right authorities and gathering information to defend one self. Spam has the tendency to crash the network by overloading it. Service providers have to buy more bandwidth, so that service to the end users can be delivered. Also as spam messages come in great bulk each day, time spent in deleting those messages is also a cost.

Table 1: Emperical survey of the taxonomy

	Value	Mean	A.L
Damage (Cronbach Alpha .73), overall mean value 2.64			
Virus can damage computer hardware components?	13.08**	2.51	P.A
Due to worms information can be enclosed to unauthorized users, it can slow down network and backdoor installation is possible.	14.86**	2.94	P.A
Trojans can open network ports and can help in carrying out denial of service attack.	13.08**	2.51	P.A
Phishing e-mails are the cause of identity theft and effetcs online business.	14.15**	2.78	P.A
Spam emails can overload CPU, freeze system and can fill up the disk space.	12.95**	2.47	P.A
Cost (Cronbach Alpha .78), overall mean value 2.49			
The cost of damages due to virus can range from business loss, information loss, time and money lost.	18.14**	3.48	M.A
To stop the worm from spreading network should be shut down this will r esult in no work for many days and can cost companies great loss.	11.84**	2.08	P.A
Service providers also faces phishing email damage cost when they have to freeze accounts, provide customer service and rest passwords.	11.84**	2.08	P.A
Users are also related to damage cost due to phishing emails in the form of tracking down the culprit, time and money spent to get identity back.	13.22**	2.55	P.A
Spam related damage cost are buying more bandwidth, financial fraud and deleting spam messages	12.34**	2.28	P.A
Propogation(Cronbach Alpha.65), overall mean 2.86			
Virus propagation can be possible through hard disk, floppy disk, files and programs.	15.45**	3.05	M.A
Virus can spread through e-mails and instant message services?	19.40**	3.64	M.A
Worms look for weaknesses in the system for the purpose of spreading without any user interaction?	12.45**	2.32	P.A
Trojan and phishing e-mails do not posses the capability of spreading but other harmful programs could be installed through them.	11.42**	1.85	L.A
Spam means of spreading is email attachments	17.85**	3.44	M.A
Precaution against Virus, worm , trojan (Cronbach Alpha .78), overall mean 3.32			
Up-to-date antivirus with patches	23.72**	4.02	H.A
Avoid using pirated software	14.86**	2.94	P.A
Avoid file sharing with unknown people	16.32**	3.21	M.A
Installing and maninting a firewall	18.14**	3.48	M.A
Do not open any suspicious emails and attachments	22.68**	3.95	M.A
When browsing websites and forums avoid clicking on advertisements	17.05**	3.33	M.A
To protect against worms do not use software which the worm exploits and fix vulnerabilities in the system.	16.32**	3.21	M.A
In case a Trojan infects system disconnect from internet to protect the confidential files.	12.82**	2.43	P.A
Precautions against Phishing (Cronbach Alpha .65) overall mean 3.01			
Check the reputation of the company when buying online.	15.24**	3.01	M.A
Take proper precautions when giving out credit cards numbers or bank details.	31.62**	4.41	H.A
Use phish blocker software	14.32**	2.82	P.A
Common precaution in Spam and Phishing (Cronbach Alpha .74), overall mean 3.04			
Never respond to phishing or spam messages	20.89**	3.79	M.A
Be careful in entering personal info on websites and forums	26.95**	4.22	H.A
Avoid opening phishing or spam e-mail attachments	23.19**	3.99	M.A
Check privacy policy on forums when subscribing	16.55**	3.25	M.A
Do not click on advertisement	12.03**	2.16	P.A
Have multiple email address	12.69**	2.39	P.A
Check URL of the website	11.93**	2.12	P.A
Report to right authorities	12.45**	2.32	P.A

C. Third aspect

Virus can be transferred form one system to another through hard disk or files and programs. For example, the virus could be present in the hard disk or any file and when these files are transferred to other computers, the virus transfers as well. On network, virus can spread when downloading from the internet or a virus can reside in an e-mail attachment. Moreover, virus can propagate when sharing files with others on the internet. Worm propagation is different from virus propagation because some types of worms usually look for weaknesses in the system. Worms are mostly written for those vulnerabilities which the end user is not aware of. Worm sends copies of itself to different computers using the network and attaches itself to addresses presented in address book. Trojans do not have the ability to copy themselves nor can they spread. Once they are installed in the system they only harm that specific system. But trojans can install harmful programs such as virus or worm, and they will propagate according to their propagation method. In phishing no propagation is noticed. This means that in case a user gets in contact with an e-mail, that e-mail will not spread to others. Phishing e-mails are usually one to one correspondence. Some phishing e-mails may have trojans or other malicious programs such as key loggers or virus and worm. These malicious programs will spread according to their propagation scheme. E-mail attachments are the number one cause of propagation because nearly every one in some manner uses e-mail. Spam can propagate through e-mail attachments. For example, an end user gets an e-mail from a friend about certain website giving good deals on products. On opening the website, the e-mail is sent to every one in the address book of that end user. In a few days the end user receives the same e-mail from other friends. This process keeps going on and the propagation will never stop until spam protection is utilized [9] [10].

D. Fourth Aspect

In order to avoid worms, system weaknesses should be fixed and those specific software's should be avoided which the worm can utilize. Some common precautions can be taken in order to avoid malware attacks. In virus, worms and trojans some common precaution are an up-to-date operating system and antivirus program. Taking safety measure when browsing the internet or checking e-mail or sharing files with others. Always take backup of files, reporting to right authorities so that the matter could be resolved and by providing feedback attacks can be avoided. In case of phishing never give out credit card numbers, bank details, always check whether the company is genuine and try using phish blocker to avoid getting such emails. To protect from spam never purchase from spam messages and always use the spam filtering option. Spam and phishing also have some common defense measures such as, never respond to phishing or spam messages, check privacy policy on forums when subscribing, have multiple e-mail addresses, be careful in entering personal information on websites and forums.

V. Conclusion

The discovery of computers have entered the man kind from old age to the new technological era. Today's rapid technological development has not only facilitated the consumers/users but at the same time has created several challenges both for computer experts as well as the end users. The expert users have developed multiple techniques to safe guard themselves from the serious ever growing threat of computer attacks but on the other end has left the end users at the mercy of so called anti-virus programs. Previously studies are more concentrated towards the development of those taxonomies that could help only the expert users in order to cope against these attacks. These taxonomies are used for a better understanding of the real problem and thus finding an appropriate solution. Therefore, the current research fulfills the gap and presents taxonomy that would prove to be beneficial for end users in understanding and diagnosing the problems caused by these serious threats and finding immediate remedies to avoid heavy costs of destruction. This taxonomy contributes to the literature and opens new avenues for future research in securing the end users, thus providing the computer users a safe heaven where they can fell secure and confident.

References

[1] R. Bisbey, and D. Hollingworth, “Protection Analysis: Final report (PA),” Technical Report ISI/RR-78-13, USC/Information Sciences Institute, May 1978.

[2] M. Bishop, "A Taxonomy of UNIX System and Network Vulnerabilities," Technical Report CSE-95-10, Univ. of California, Sept. 1995.

[3] C.E. Landwehr, A.R. Bull, J.P. McDermott and W.S. Choi, “A Taxonomy of Computer Program Security Flaws,” ACM Computing Surveys, vol. 26, no. 3, pp. 211–254, Sept. 1994.

[4] J.D. Howard, “An Analysis of Security Incidents on the Internet, 1989-1995,” PhD thesis, Dept. of Eng. and Public Policy, Carnegie-Mellon Univ., Apr. 1997.

[5] S. Hansman, R. Hunt, "A Taxonomy of network and computer attacks," Computers & Security, vol. 24, pp. 31-43, 2005.

[6] F-Secure IT Security Threat Summary for the Second Half of 2008. Avaiable: http://www.f-secure.com/en_EMEA-Labs/news-info/threat-summaries/2008/2008-4.html

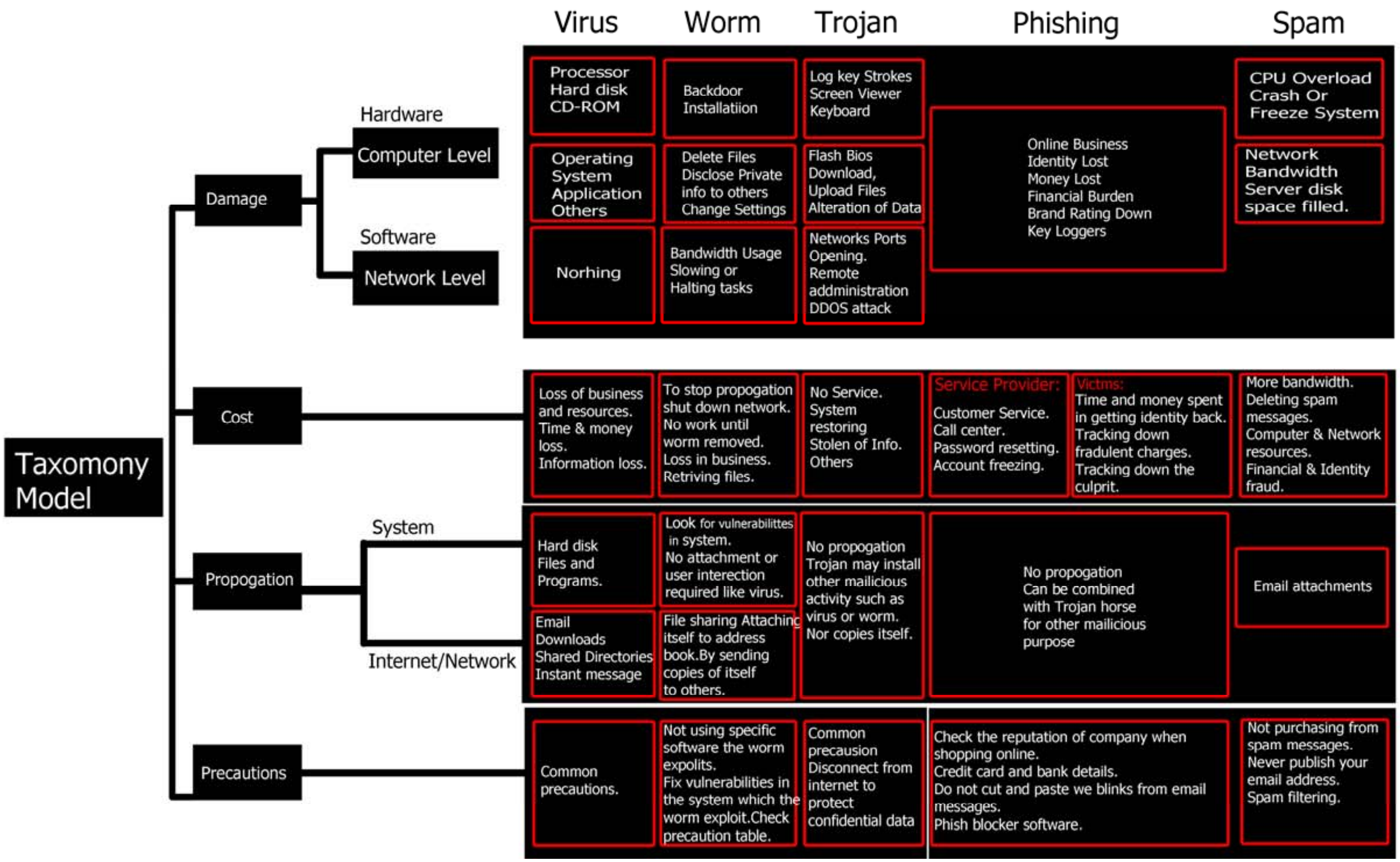
[7] CERT statistics Software engineering institute Carnegie Mellon University, Feburary 2009. Avaliable : www.cert.org/stats/cert_stats.html; 2009.

[8] E. Udassin, “Control system attack vectors and example : Field Site and Corporate Network” SCADA Security Scientific Symposium, 2008.

[9] W. Stallings, Network Security Essentials applications and standards. Upper Saddle River, New Jersey: Prentice Hall 2007.pp. 332-348

[10] D. Salomon. Foundations of Computer Security. London: Springer-Verlag 2006. pp 43, 66, 91, 113, 169

TABLE 2: MALICIOUS PROGRAM TAXONOMY FOR END USER



Taxomony Model			Virus	Worm	Trojan	Phishing	Spam
Damage	Hardware	Computer Level	Processor Hard disk CD-ROM	Backdoor Installatiion	Log key Strokes Screen Viewer Keyboard	Online Business Identity Lost Money Lost Financial Burden Brand Rating Down Key Loggers	CPU Overload Crash Or Freeze System
			Operating System Application Others	Delete Files Disclose Private info to others Change Settings	Flash Bios Download, Upload Files Alteration of Data		Network Bandwidth Server disk space filled.
	Software	Network Level	Norhing	Bandwidth Usage Slowing or Halting tasks	Networks Ports Opening. Remote addministration DDOS attack		
Cost			Loss of business and resources. Time & money loss. Information loss.	To stop propogation shut down network. No work until worm removed. Loss in business. Retriving files.	No Service. System restoring Stolen of Info. Others	Service Provider: Customer Service. Call center. Password resetting. Account freezing. Victms: Time and money spent in getting identity back. Tracking down fradulent charges. Tracking down the culprit.	More bandwidth. Deleting spam messages. Computer & Network resources. Financial & Identity fraud.
Propogation	System		Hard disk Files and Programs.	Look for vulnerabilittes in system. No attachment or user interection required like virus.	No propogation Trojan may install other mailicious activity such as virus or worm. Nor copies itself.	No propogation Can be combined with Trojan horse for other mailicious purpose	Email attachments
	Internet/Network		Email Downloads Shared Directories Instant message	File sharing Attaching itself to address book.By sending copies of itself to others.			
Precautions			Common precautions.	Not using specific software the worm expolits. Fix vulnerabilities in the system which the worm exploit.Check precaution table.	Common precausion Disconnect from internet to protect confidential data	Check the reputation of company when shopping online. Credit card and bank details. Do not cut and paste we blinks from email messages. Phish blocker software.	Not purchasing from spam messages. Never publish your email address. Spam filtering.

Visualization of MUSTAS Model using ECHAID

G.Paul Suthan
Head, Department of Computer Science
CSI Bishop Appasamy College
Race Course, Coimbatore,
Tamil Nadu 641018, India
gpsuthan@hotmail.com

Lt.Dr.Santosh Baboo
Reader,PG and Research Department of Computer Application
DG Vishnav College, Arumbakkam
Chennai 600106,Tamil Nadu,India
Santos2001@sify.com

***Abstract*— Educational assessment is an important insight to know about the student. In recent years there is an increasing interest of Educational Data Mining (EDM), which helps to explore the student data in different perspective. As the case, we introduced a new model called MUSTAS to assess the student's attitude in three dimensions known as self assessment, institutional assessment and external assessment. Thus, this model exhibits the student performance in three grades as poor, fair, and good. The final part of visualization is generated through ECHAID algorithm. In this paper, we present the model and its performance on our private student dataset collected by us. Our model shows interesting insights about the student and can be used to identify their performance grade.**

Keywords-component; Educational Data Mining, MUSTAS, CHAID prediction, Latent Class Analysis, Hybrid CHAID, ECHAID

I. INTRODUCTION

In the past years, researchers from varity of disciplines(including computer science, statistics , data mining , and education) have started to investigate how we can improve education using Data mining concepts. As a result Educational Data Mining[EDM] has emerged. EDM emphasis on developing methods on exploring unique type of data that come from educational context. Educational Data Mining is concerned with developing methods for exploring data from educational settings. Data mining also called Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large amount of data by Witten and Frank[19]. It has been proposed that educational data mining methods are often different from standard data mining methods, due to the need to explicitly account for educational data by Baker[3]. For this reason, it is increasingly common to see the use of models in these series as suggested by Barnes[4] and Pavlik et al.[14]. The traditional data mining methods are constructed in generic pattern, which is suitable for any kind of application to fit in the method specified. Hence, existing techniques may be useful to discover the data, but it does not fulfill specific or customized requirement.

Education specific mining techniques can help to improve the instructional design, understanding of student's attitude, academic performance appraisal and so on. In this scenario, traditional mining algorithms need to be adjusted into educational context.

II. RESEARCH BACKGROUND

Modern educational and psychological assessment is dominated by two mathematical models, Factor Analysis (FA) and Item Response Theory (IRT). FA operates at the level of a test, i.e., a collection of questions (items). The basic assumption of FA is that test score of individual i on test j is determined by

$$x_{ij} = \sum_{k=1}^{K} w_{kj} f_{ik} + e_{ij} \qquad (1)$$

where the f_{ik} terms represent the extent to which individual i has underlying ability k, and the w_{kj} terms represent the extent to which the ability k is required for test j. The e_{ij} term is a residual which is to be minimized. The weights of the abilities required for the test, i.e. the $\{w_{kj}\}$, is constant across individuals. This amounts to an assumption that all individuals deploy their abilities in the same way on each test. Assessments are made in an attempt to determine students' f_{ik} values, i.e. a student's place or position on the underlying ability scales.

IRT operates at the item level within a test. Consider the i^{th} item on a test. This item is assumed to have a characteristic difficulty level, B_i. Each examinee is assumed to have skill level θ on the same scale. In the basic three parameter IRT model, the probability that a person with ability θ will get item i correct is

$$P(\theta) = c_i + (1\text{-}c_i)\frac{e^{Da_i(\theta - B_i)}}{1 + e^{Da_i(\theta - B_i)}} \qquad (2)$$

where D is a constant scaling factor, a_i is an item discrimination parameter and c_i is a "correction for guessing parameter". A consequence of this model is that the relative order of difficulty for any pair of items on a test and must be the same for all individuals.

Neither any of these commonly used models allow for idiosyncratic patterns of thought, where different people attack problems in different ways. More specialized models can describe mixtures of strategies as mentioned by Huang[9]. However, many educational theories are not easily fit to the assumptions of factor analytic or IRT models. Much of the motivation behind diagnostic assessment is to identify the different strategies that might change the relative order of difficulty of items.

The problem of how best to mathematically model a knowledge space is open, and the answer may be domain-dependent. There is evidence suggesting that in fact, facets (fine grained correct, partially correct, and incorrect understandings) may have a structure to them in some domains that can be modeled using a partial credit model as described by Wright and Masters[22]. Using this model, multiple choice responses are ordered in difficulty on a linear scale, allowing one to rank students by ability based on their responses. This implies that the relative difficulty of items in some interesting domains may indeed be the same for all students as said by Scalise et al.[17]. Thus, modeling can be improved by identifying this linear structure of concepts. Wilson[20] and Wislon and Sloane[21] mentions that each item response would have its own difficulty on a linear scale, providing a clear measure of student and classroom progress, e.g., a learning progression where content is mapped to an underlying continuum. But building this knowledge representation is an extremely large endeavor, especially in subject areas where little research has been done into the ideas students have before instruction that affect their understanding, or what dimensional structure is appropriate to represent them. This approach assumes that all options are equally plausible, because if one option made no sense, even the lowest ability person would be able to discard it, so IRT parameter estimation methods take this into account and estimate a c_i based on the observed data. In contrast, automatically constructed knowledge spaces may lead to overestimation of knowledge states. Thus, we have paid attention to create a unique framework based on exploratory data-mining approach.

III. EDUCATIONAL DATA MINING (EDM)

In recent years, advances in computing and information technologies have radically expanded the data available to researchers and professionals in a wide variety of domains. EDM has emerged over past few years, and its community has actively engaged in creating large repositories. The increase in instrumented educational software and in databases of student test scores has created large data repositories reflecting how students learn. EDM focuses on computational approaches for using those data to address important educational questions.

Erdogan and Timor [5] used educational data mining to identify and enhance educational process which can improve their decision making process. Henrik[8] concluded that clustering was effective in finding hidden relationships and associations between different categories of students. Walters and Soyibo,[23] conducted a study to determine Jamaican high school students' (population n=305) level of performance on five integrated science process skills with performance linked to gender, grade level, school location, school type, student type, and socioeconomic background (SEB). The results revealed that there was a positive significant relationship between academic performance of the student and the nature of the school.

Khan, [24] conducted a performance study on 400 students comprising 200 boys and 200 girls selected from the senior secondary school of Aligarh Muslim University, Aligarh, India with a main objective to establish the prognostic value of different measures of cognition, personality and demographic variables for success at higher secondary level in science stream. The selection was based on cluster sampling technique in which the entire population of interest was divided into groups, or clusters, and a random sample of these clusters was selected for further analyses. It was found that girls with high socio-economic status had relatively higher academic achievement in science stream and boys with low socioeconomic status had relatively higher academic achievement in general.

Hijazi and Naqvi, [18] conducted a study on the student performance by selecting a sample of 300 students (225 males, 75 females) from a group of colleges affiliated to Punjab university of Pakistan. The hypothesis that was stated as "Student's attitude towards attendance in class, hours spent in study on daily basis after college, students' family income, student mother's age and mother's education are significantly related with student performance" was framed. By means of simple linear regression analysis, it was found that the factors like mother's education and student's family income were highly correlated with the student academic performance.

A.L Kristjansson, Sigfusdottir and Allegrante[2] made a study to estimate the relationship between health behaviors, body mass index (BMI), self-esteem and the academic achievement of adolescents. The authors analyzed survey data related to 6,346 adolescents in Iceland and it was found that the factors like lower BMI, physical activity, and good dietary habits were well associated with higher academic achievement. Therefore the identified students were recommended diet to suit their needs.

Cortez and Silva[15] attempted to predict failure in the two core classes (Mathematics and Portuguese) of two secondary school students from the Alentejo region of Portugal by utilizing 29 predictive variables. Four data mining algorithms such as Decision Tree (DT), Random Forest (RF), Neural Network (NN) and Support Vector Machine (SVM) were applied on a data set of 788 students, who appeared in 2006 examination. It was reported that DT and NN algorithms had the predictive accuracy of 93% and 91% for two-class dataset (pass/fail) respectively. It was also reported that both DT and

NN algorithms had the predictive accuracy of 72% for a four-class dataset.

IV. MUSTAS MODEL

The Multidimensional Students Assessment (MUSTAS) framework is a novel model, which consist of demographic factors, academic performance of the student and dimensional factors. The dimensional factors has further sub divided into three dimensions respectively self assessment, institutional assessment and external assessment. The main objective of this framework is to identify the contribution of selected dimensions over academic performance of the student, which helps to teachers, parents and management about the student's pattern. Understanding of the pattern may facilitate to redefine the education method, additional care on weakness, and promoting their abilities.

A general form of the Multidimensional Random Coefficient Multinomial Logit Model was fitted, with between-item dimensionality as described by Adams, Wilson & Wang[1]. This means each item was loaded on a single latent dimension only so that different dimensions contained different items. A three-dimensional model, a two-dimensional model and a one-dimensional model were fitted in sequence. The three-dimensional model assigned items into three groups. Group 1 consisted of items that had a heavy reading and extracting information component. Group 2 consisted of items that were essentially common-sense mathematics, or non-school mathematics. Group 3 consisted of the rest of the item pool, consisting of mostly items that were typically school mathematics, as well as logical reasoning items. In this item response theory (IRT) model, Dimensions 3 and 4 of the framework, mathematics concepts and computation skills, had been combined to form one IRT dimension.

The MUSTAS model was built with the backbone of CHAID and LCM. Chi-squared Automatic Interaction Detection (CHAID) analysis which was first proposed by Kass, 1980[6] is one of post-hoc predictive segmentation methods. The CHAID, using of decision tree algorithms, is an exploratory method for segmenting a population into two or more exclusive and exhaustive subgroups by maximizing the significance of the chi-square, based on categories of the best predictor of the dependent variable. Segments obtained from CHAID analysis are different from cluster type models because the CHAID method, which is derived to be predictive of a criterion variable, is defined by combinations of predictor variables by Magidson, [12].

Latent Class (LC) modeling was initially introduced by Lazarsfeld and Henry.[10] as a way of formulating latent attitudinal variables from dichotomous survey items. In contrast to factor analysis, which posts continuous latent variables, LC models assume that the latent variable is categorical, and areas of application are more wide ranging. In recent years, LC models have been extended to include observable variables of mixed scale type (nominal, ordinal, continuous and counts), covariates, and to deal with sparse data, boundary solutions, and other problem areas.

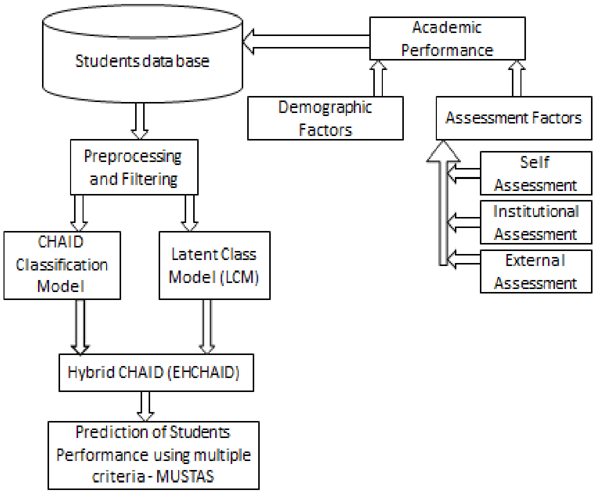


Figure 1: MUSTAS Model

The Figure 1, exhibits the proposed model of student assessment strategy. Academic performance and assessment factors are combined together as General Assessment Classification (GAC), which is visualize through demographic factors of the students. The GAC can be mentioned as parameter, which is act as rule based classification.

AMOS is an application for structural equation modeling, multi-level structural equation modeling, non-linear modeling, generalized linear modeling and can be used to fit measurement models to data. In the subsequent sections, we illustrate this feature by fitting a measurement model to an SPSS data set using path diagram.

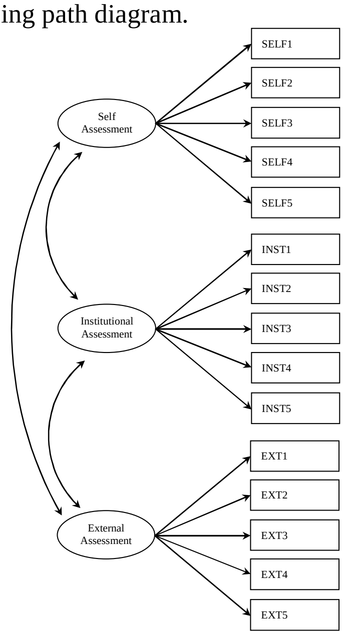


Figure 2. Path Diagram of MUSTAS

The path diagram shown in Figure 2, exhibit the pattern of MUSTAS model, which extracts R2=0.802. The LC analysis used to identifying segments based on academic performance of the student and observed assessment score into K underlying latent class segments.

V. ECHAID ALGORITHM

The ECHAID algorithm is a hybrid methodology combining features of CHAID and latent class modeling (LCM) to build a classification tree that is predictive of multiple criteria. This algorithm is derived from Hybrid CHAID algorithm and it involves four steps.

1. Abstract the specified factors into *D* dimensional variables.
2. Perform an LC cluster analysis on *D* response variables to obtain *K* latent classes.
3. Perform a CHAID analysis using the K classes as nominal dependent variable.
4. Obtain predictions for each of D dimensional variables based on resulting CHAID segments and/or on any preliminary set of CHAID segments.

Step 1 yields an abstraction of specified number of factors into dimension. Similarly it is recalled for number of dimensions required.

$$D = \frac{\sum_{n=1}^{N} cw}{\sum c} \qquad (3)$$

The variable *D* is a dimension, *c* is factor(s) which is proposed for abstraction and *w* is weight. The rounded value of *D* is equal to scale. Hence this step optimizes the number of factors into dimensional variable with similar scale.

Step 2 yields class-specific predicted probabilities for each category of the d-th dependent variable, as well as posterior membership probabilities for each case.

$$\begin{aligned} P(\mathbf{Y}=\mathbf{j}) &= \sum_{k=1}^{K} P(X=k,\mathbf{Y}=\mathbf{j}) = \sum_{k=1}^{K} P(X=k)P(\mathbf{Y}=\mathbf{j}|X=k) \\ &= \sum_{k=1}^{K} P(X=k) \prod_{d=1}^{D} P(Y_d = j_d|X=k), \end{aligned} \qquad (4)$$

Step 3 yields a set of CHAID segments that differ with respect to their average posterior membership probabilities for each class. We use the posterior membership probabilities defined in equation as fixed case weights as opposed to the modal assignment into one of the *K* classes. This weighting eliminates bias due to the misclassification error that occurs if cases were equated (with probability one) to that segment having the highest posterior probability. Specifically, each case contributes *K* records to the data, the kth record of which contains the value *k* for the dependent variable, and contains a case weight of P(X = *k*|Y = *j*), the posterior membership probability associated with that case. Thus, as opposed to the original algorithm where chi-square is calculated on observed 2-way tables, in the hybrid algorithm, the chi-squared statistic is computed on 2-way tables of weighted cell counts.

If as an alternative to performing a standard LC analysis, one performs an LC factor analysis in step 1, in step 2 the CHAID ordinal algorithm can be used to obtain segments based on the use of any of the LC factors as the ordinal dependent variable, or a single segmentation can be obtained using the nominal algorithm to identify segments based on the single joint latent variable defined as a combination of two or more identified LC factors.

$$P(X=k|\mathbf{Y}=\mathbf{j}) = \frac{P(X=k,\mathbf{Y}=\mathbf{j})}{P(\mathbf{Y}=\mathbf{j})}. \qquad (5)$$

Step 4 involves obtaining predictions for any or all of the *D* dependent variables for each of the *I* CHAID segments by cross-tabulating the resulting CHAID segments by the desired dependent variable(s). An alternative is to obtain predictions as follows

$$P(Y_d = j|i) = \sum_{k=1}^{K} P(Y_d = j|X=k)P(X=k|i). \quad (6)$$

As can be seen, we compute a weighted average of the class-specific distributions for dependent variable Y_d obtained in step 2 *[P(Y$_d$ = j|X = k)]*, with the average posterior membership probabilities obtained in step 3 for segment *i* being used as the weights *[P(X = k|i)]*.

VI. RESULTS AND DISCUSSION

Dataset was prepared based on the feedback collected from students of various colleges in Coimbatore city. The overall data finalized for dataset was 1000, which is inclusive of errors. After preprocessing the final dataset was optimized to 933 through the elimination of nonresponsive or error records. The main intention of this paper is to prove the MUSTAS framework developed by Paul suthan and Santhosh Baboo[13] is perfectly fit into Hybrid CHAID algorithm proposed by Magidson and Vermunt[11]. In their proposal they have considered multiple criteria, whereas in MUSTAS framework multiple dimensions are considered such as Self Assessment, Institutional Assessment and External Assessment. Each dimension possessing five factors. Hence we proposed an abstract layer on top of Hybrid CHAID, which is performing optimization of factors into respective dimension with similar scale. The results are generated using Latent GOLD and SI-CHAID software founded by Statistical Innovations Inc. Latent GOLD is a powerful latent class modeling software, which supports three different model structures such as cluster models, discrete factor (DFactor) models and regression models. SI-CHAID is software, which is inter-related with

Latent GOLD to perform CHAID tree visualization. A LCM was fit to these data, using college academic performance as an active covariate and 8 demographics as inactive covariates along with academic performance (AP) + three dimensions (SA, IA, EA) which are considered as multiple dependent variables. By default the LCM yielded 3 segments, which is similar to Hybrid CHAID. The first segment (class 1) 22% Good performers, second segment (class 2) 58.5% Moderate performers and third segment (class 3) 19.5% poor performer, which are predicted from dimensional rating and academic performance. It is presented in Fig.3 root node.

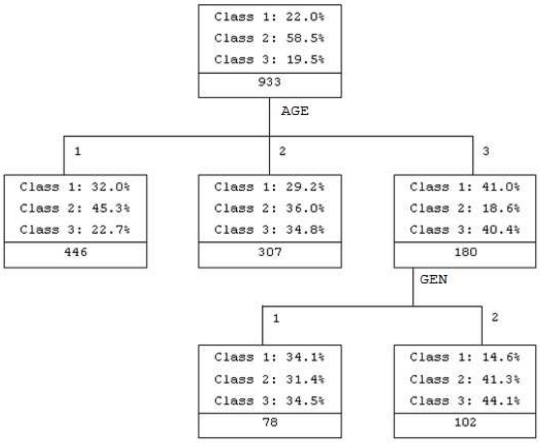


Figure 3. ECHAID Tree for 4 Dependent Variables

Tree visualization shown in this paper is based on the default parameter. Hence ECHAID used the three classifications as dependent variables and eight demographic variables as the predictors. The Figure 3, states that at root node two predictors (AGE, GEN) out of eight predictors were found significant, which is less than 0.001. The inner classification of each node depicts the distribution pattern of the students in the respective classification. Fig. 4 depicts the hybrid segments to predict academic performance, which is trying to assess good performer.

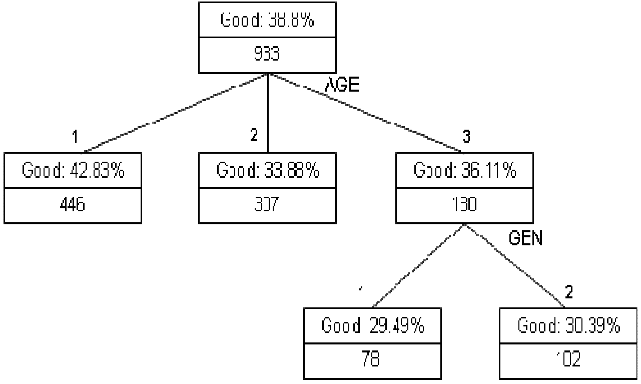


Figure 4. EHCHAID **tree – Good Performer**

The Figure 4, shows that among total number of students considered for this evaluation 38.8% of them are good performers. The age has three categories below 19 years, 19-21 years and above 21 years. Similarly gender has two categories male and female. It is noticed that 446 students fall under below 19 years age group, out of which 42.83% of them are good performer, 33.88% under 19-21 years age group and 36.11% of them fall under above 21 years age group. Among 180 students under above 21 years age group, 78 students were male and good performers are 29.49% and 102 students were female and good performers are 30.39%. Hence the ECHAID tree helps to assess the student performance and it proves that MUSTAS framework is perfectly suitable for Educational Data Mining purpose and closely associated with Hybrid CHAID.

VII. CONCLUSION

In this paper, we introduced ECHAID algorithm as a model for MUSTAS framework, which supports multiple dependent variables. It enhances visualization of traditional CHAID algorithm and provides unique segmentations. In future work, performance analysis, feature extraction and classification accuracy can be evaluated.

REFERENCES

[1] Adams, R. J., Wilson, M R. & Wang, (1997), "The M. multidimensional random coefficients multinomial logit model", Applied Psychological Measurement, 21, 1-233.

[2] A. L. Kristjansson, I. G. Sigfusdottir, and J. P. Allegrante, "Health Behavior and Academic Achievement among Adolescents: The Relative contribution of Dietary Habits, Physical Activity, Body Mass Index, and Self-Esteem", Health Education &Behavior, (In Press).

[3] Baker, R.S.J.D., Barnes, T. and Beck, (2008), 1st International Conference on Educational Data Mining, Montreal, Quebec, Canada.

[4] Barnes.T, (2005), "The q-matrix method: Mining student response data for knowledge", In proceedings of the AAAI-2005 Workshop on Educational Data mining.

[5] Erdogan and Timor, (2005), "A data mining application in a student database", Journal of Aeronautic and Space Technologies July 2005 Vol. 2 No. 2 (53-57).

[6] G.V. Kass, (1980), "An Exploratory Technique for Investigating Large Quantities of Categorical Data", Applied Statistic, Vol. 29, pp. 119-127.

[7] S. T. Hijazi, and R. S. M. M. Naqvi, "Factors Affecting Student's Performance: A Case of Private Colleges", Bangladesh e-Journal of Sociology, Vol. 3, No. 1, 2006.

[8] Henrik (2001), "Clustering as a Data Mining Method in a Web-based System for Thoracic Surgery".

[9] Huang, (2003), "Psychometric Analysis Based on Evidence-Centered Design and Cognitive Science of Learning to Explore Student's Problem-Solving in Physics", University of Maryland.

[10] Lazarsfeld,P.F., and Henry, (1968), "Latent Structure Analysis", Boston: Houghton Mill.

[11] Madigson J and Jeroen Vermunt "An Extension of the CHAID Tree-based Segmentation Algorithm to Multiple Dependent Variables" (1) Statistical Innovations Inc., 375 Concord Avenue, Belmont, MA 02478, USA (2) Department of Methodology and Statistics, Tilburg University, PO Box 90153,5000 LE Tilburg, Netherlands

[12] Magidson.J,(1994), “The CHAID approach to segmentation modeling: Chi-squared automatic interaction detection”, In advanced methods of marketing research, Cambridge, Blackwell, pp. 118-159.

[13] Paul Suthan. G and Santhosh Baboo(2011),” Hybrid CHAID a key for MUSTAS Framework in Educational Data Mining” IJCSI International Journal for computer Science Issues, Vol8 , Issue 1, January 2011.

[14] Pavlik.P., Cen, H., Wu, L. and Koedinger.K, (2008), “Using Item-type Performance Covariance to Improve the Skill Model of an Existing Tutor”, In Proceedings of the 1st International Conference on Educational Data Mining, pp. 77-86.

[15] P.Cortez and A.Silva, (2008), “Using Data Mining to Predict Secondary School Student Performance”, In EUROSIS, pp.5-12.

[16] Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M. Inayat Khan, (2010), “Data mining model for higher education system”, European Journal of Scientific Research, Vol.43, No.1, pp.24-29.

[17] Scalise, K., Madhyastha, T., Minstrell, J. and Wilson, M.,(in-press), “Improving Assessment Evidence in e-Learning Products: Some Solutions for Reliability”, International Journal of Learning Technology (IJLT).

[18] S. T. Hijazi, and R. S. M. M. Naqvi, (2006), “Factors Affecting Student’s Performance: A Case of Private Colleges”, Bangladesh e-Journal of Sociology, Vol. 3, No. 1.

[19] Witten, I.H. and Frank.E, (1999), “Data mining: Practical Machine Learning Tools and Techniques with Java Implementations”, Morgan Kaufmann, San Fransisco, CA.

[20] Wilson.M, (2004), “Constructing Measures: An Item Response Modeling Approach”, Lawrence Erlbaum.

[21] Wilson.M and Sloane.K, (2000), “From Principles to Practice: An Embedded Assessment System”, Applied Measurement in Education 13.

[22] Wright, B.D. and Masters.G.N, (1982), “Rating Scale Analysis”, Pluribus.

[23] Y. B. Walters, and K. Soyibo, (2001), “An Analysis of High School Students' Performance on Five Integrated Science Process Skills”, Research in Science & Technical Education, Vol. 19, No. 2, pp.133-145.

[24] Z. N. Khan, “Scholastic Achievement of Higher Secondary Students in Science Stream”, Journal of Social Sciences, Vol. 1, No. 2, 2005, pp84-87.G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

AUTHORS PROFILE

G. Paul Suthan has done his Under-Graduation and Post-Graduation at Bishop Heber College, affiliated to Bharathidasan University and Master of Philosophy at Manonmaniam Sundaranar University. He is currently pursuing his Ph.D in Computer Science in Dravidian University, Kuppam, Andhra Pradesh. Also, he is working as the Head of the Department of MCA, Bishop Appasamy College of Arts and Science, Coimbatore, affiliated to Bharathiar University. He has organized various National and State level seminars, and Technical Symposium. He has participated in various National conferences and presented papers. He has 15 years of teaching experience. His research areas include Data Mining and Artificial Intelligence.

Lt.Dr.S.Santhosh Baboo, aged forty three, has around twenty years of postgraduate teaching experience in Computer Science, which includes Seven years of administrative experience. He is a member, board of studies, in several autonomous colleges, and designs the curriculum of undergraduate and postgraduate programmes. He is a consultant for starting new courses, setting up computer labs, and recruiting lecturers for many colleges. Equipped with a Masters degree in Computer Science and a Doctorate in Computer Science, he is a visiting faculty to IT companies. It is customary to see him at several national/international conferences and training programmes, both as a participant and as a resource person. He has been keenly involved in organizing training programmes for students and faculty members. His good rapport with the IT companies has been instrumental in on/off campus interviews, and has helped the post graduate students to get real time projects. He has also guided many such live projects. Lt.Dr. Santhosh Baboo has authored a commendable number of research papers in international/national Conference/journals and also guides research scholars in Computer Science. Currently he is Reader in the Postgraduate and Research department of Computer Science at Dwaraka Doss Goverdhan Doss Vaishnav College (accredited at ‘A’ grade by NAAC), one of the premier institutions in Chennai.Authors Profile …

Optimized Energy and QoS Aware Multi-path Routing Protocol in Wireless Sensor Networks

Mohammad Reza Mazaheri
Department of Technical and Engineering
Mashhad Branch , Islamic Azad University
Mashhad, Iran
Mohammad.Mazaheri1@gmail.com

Sayyed Majid Mazinani
Department of Electrical Engineering
Imam Reza University
Mashhad, Iran
Mazinani@ieee.org

***Abstract*—Satisfying Quality of Service (QoS) requirements (e.g. bandwidth and delay constraints) for the different QoS based applications of WSNs raises significant challenges. Each algorithm that is used for packet routing in such applications should be able to establish tradeoffs between end to end delay parameter and energy consumption. Therefore, enabling QoS applications in sensor networks requires energy and QoS awareness in different layers of the protocol stack. In this paper, we propose an Optimized Energy and QoS Aware Multipath routing protocol in wireless sensor networks namely OEQM. This protocol maximizes the network lifetime via data transmission across multiple paths as load balancing that causes energy consume uniformly throughout the network. OEQM uses the residual energy, available buffer size, Signal-to-Noise Ratio (SNR) and distance to sink to predict the best next hop through the paths construction phase also our protocol employs a queuing model to handle both real-time and non-real-time traffic. Simulation results show that our proposed protocol is more efficient than previous algorithms in providing QoS requirements and minimizing energy consumption.**

Keywords-multi-path; network lifetime; energy consumption; Qos requirements; cost metric

I. INTRODUCTION

In the recent years, the rapid advances in micro-electromechanical systems, low power and highly integrated digital electronics, small scale energy supplies, tiny microprocessors and low power radio technologies have created low power, low cost and multifunctional wireless sensor devices, which can observe and react to changes in physical phenomena of their environments. These sensor devices are equipped with a small battery, a tiny microprocessor, a radio transceiver and a set of transducers that used to gathering information that report the changes in the environment of the sensor node. The emergence of these low cost and small size wireless sensor devices has motivated intensive research in the last decade addressing the potential of collaboration among sensors in data gathering and processing, which led to the creation of Wireless Sensor Networks (WSNs) .

A typical WSN consists of a number of sensor devices that collaborate with each other to accomplish a common task (e.g. environment monitoring, object tracking, etc.) and report the collected data through wireless interface to a sink node. The areas of applications of WSNs vary from civil, healthcare and environmental to military. Examples of applications include target tracking in battlefields, habitat monitoring, civil structure monitoring, forest fire detection and factory maintenance [1].

However, with the specific consideration of the unique properties of sensor networks such limited power, stringent bandwidth, dynamic topology (due to nodes failures or even physical mobility), high network density and large scale deployments have caused many challenges in the design and management of sensor networks. These challenges have demanded energy awareness and robust protocol designs at all layers of the networking protocol stack [2].

Efficient utilization of sensor’s energy resources and maximizing the network lifetime were and still are the main design considerations for the most proposed protocols and algorithms for sensor networks and have dominated most of the research in this area. However, depending on the type of application, the generated sensory data normally have different attributes, where it may contain delay sensitive and reliability demanding data. Furthermore, the introduction of multimedia sensor networks along with the increasing interest in real time applications have made strict constraints on both throughput and delay in order to report the time-critical data to the sink within certain time limits and bandwidth requirements without any loss. These performance metrics (i.e. delay and bandwidth) are usually referred to as Quality of Service (QoS) requirements [3]. Therefore, enabling many applications in sensor networks requires energy and QoS awareness in different layers of the protocol stack in order to have efficient utilization of the network resources and effective access to sensors readings. Authors of [3] and [4] have surveyed the QoS based routing protocol in WSNs.

Many routing solutions specifically designed for WSNs have been proposed in [5] and [6]. In these proposals, the unique properties of the WSNs have been taken into account. These routing techniques can be classified according to the protocol operation into negotiation based, query based, QoS based and multi-path based. The negotiation based protocols have the objective to eliminate the redundant data by include high level data descriptors in the message exchange. In query based protocols, the sink node initiates the communication by broadcasting a query for data over the network. The QoS based protocols allow sensor nodes to make tradeoffs between the energy consumption and some QoS metrics before delivering the data to the sink node [7]. Finally, multi-path routing protocols use multiple paths rather than a single path in order to improve the network performance in terms of reliability and robustness. Multi-path routing establishes multiple paths between the source-destination pair. Multi-path routing protocols have been discussed in the literature for several years now [8]. Multi-path routing has focused on the use of multiple paths primarily for load balancing, fault tolerance, bandwidth aggregation and reduced delay. We focus to guarantee the required quality of service through multi-path routing.

The rest of the paper organized as follows: in section 2, we explain some of the related works. Section 3 describes the proposed protocol with detailed. Section 4 presents the performance evaluation. Finally, we conclude the paper in Section 5.

II. Related Works

QoS-based routing in sensor networks is a challenging problem because of the scarce resources of a sensor node. Thus, this problem has received a significant attention from the research community, where many works are being made. In this section we do not give a comprehensive summary of the related work, instead we present and discuss some works related to the proposed protocol.

One of the early proposed routing protocols that provide some QoS is the Sequential Assignment Routing (SAR) protocol [9]. SAR protocol is a multi-path routing protocol that makes routing decisions based on three factors: energy resources, QoS on each path and packet's priority level. Multiple paths are created by building a tree rooted at the source to the destination. During construction of paths those nodes which have low QoS and low residual energy are avoided. Upon the construction of the tree most of the nodes will belong to multiple paths. To transmit data to sink, SAR computes a weighted QoS metric as a product of the additive QoS metric and a weighted coefficient associated with the priority level of the packet to select a path. Employing multiple paths increases fault tolerance, but SAR protocol suffers from the overhead of maintaining routing tables and QoS metrics at each sensor node.

K. Akkaya and M. Younis in [10] proposed a cluster based QoS aware routing protocol that employs a queuing model to handle both real-time and non real time traffic. The protocol only considers the end-to-end delay. The protocol associates a cost function with each link and uses the K least-cost path algorithm to find a set of the best candidate routes. Each of the routes is checked against the end-to-end constraints and the route that satisfies the constraints is chosen to send the data to the sink. All nodes initially are assigned the same bandwidth ratio which makes constraints on other nodes which require higher bandwidth ratio. Furthermore, the transmission delay is not considered in the estimation of the end-to-end delay which sometimes results in selecting routes that do not meet the required end-to-end delay. However, the problem of bandwidth assignment is solved in [11] by assigning a different bandwidth ratio for each type of traffic for each node.

SPEED [12] is another QoS based routing protocol that provides soft real-time end-to-end guarantees. Each sensor node maintains information about its neighbours and exploits geographic forwarding to find the paths. To ensure packet delivery within the required time limits, SPEED enables the application to compute the end-to-end delay by dividing the distance to the sink by the speed of packet delivery before making any admission decision. Furthermore, SPEED can provide congestion avoidance when the network is congested. However, while SPEED has been compared with other protocols and it has showed less energy consumption than other protocols, this does not mean that SPEED is energy efficient, because the protocols used in the comparison are not energy aware. SPEED does not consider any energy metric in its routing protocol, which makes a question about its energy efficiency. Therefore, to better study the energy efficiency of the SPEED protocol; it should be compared with energy aware routing protocols.

Felemban [13] propose Multi-path and Multi-Speed Routing Protocol (MMSPEED) for probabilistic QoS guarantee in WSNs. Multiple QoS levels are provided in the timeliness domain by using different delivery speeds while various requirements are supported by probabilistic multipath forwarding in the reliability domain.

X. Huang and Y. Fang have proposed multi constrained QoS multi-path routing (MCMP) protocol [14] that uses braided routes to deliver packets to the sink node according to certain QoS requirements expressed in terms of reliability and delay. The problem of the end-to-end delay is formulated as an optimization problem and then an algorithm based on linear integer programming is applied to solve the problem. The protocol objective is to utilize the multiple paths to augment network performance with moderate energy cost. However, the protocol always routes the information over the path that includes minimum number of hops to satisfy the required QoS which leads in some cases to more energy consumption.

Authors in [15], have proposed the Energy constrained multi-path routing (ECMP) that extends the MCMP protocol by formulating the QoS routing problem as an energy optimization problem constrained by reliability playback delay and geo-spatial path selection constraints. The ECMP protocol trades between minimum number of hops and minimum energy by selecting the path that satisfies the QoS requirements and minimizes energy consumption.

Meeting QoS requirements in WSNs introduces certain overhead into routing protocols in terms of energy consumption, intensive computations and significantly large storage. This overhead is unavoidable for those applications that need certain delay and bandwidth requirements. In OEQM protocol, we combine different ideas from the previous protocols in order to optimally tackle the problem of QoS in sensor networks. In this protocol we try to satisfy the QoS requirements with the minimum energy. Our routing protocol performs routes discovery using multiple criteria such as residual energy, remaining buffer size, signal-to-noise ratio and distance to sink.

III. Description of The Proposed Protocol

In this section, we first define some assumptions, then we provide the details of multiple paths discovery and maintenance as well as the traffic allocation and data transmission across the multiple paths.

A. Assumptions

We assume *N* identical sensor nodes are distributed randomly in the sensing filed. All nodes have the same transmission range and have enough battery power to carry their sensing, computing and communication activities. The sink is not mobile and considered to be a powerful node endowed with enhanced communication and computation capabilities as well as no energy constraints. The network is fully connected and each node in the network is assigned a unique ID also all nodes are willing to participate in communication process by forwarding data. Furthermore, at any time, we assume that each sensor node is able to compute its distance to sink, its residual energy and its available buffer size (remaining memory space to cache the sensory data while it is waiting for servicing) as well as record the link performance between itself and its neighbor node in terms of signal-to noise ratio (SNR) and distance to sink.

B. Path Discovery Mechanism

In multi-path routing, node-disjoint paths (i.e. have no common nodes except the source and the destination) are usually preferred because they utilize the most available network resources, hence are the most fault-tolerant. If an intermediate node in a set of node-disjoint paths fails, only the path containing that node is affected, so there is a minimum impact to the diversity of the routes [16]. Based on the idea of the directed diffusion [17], the sink node starts the multiple paths discovery phase to create a set of neighbours that able to forward data towards the sink from the source node.

In first phase of path discovery procedure, each sensor node broadcast a HELLO message to its neighbouring nodes in order to have enough information about which of its neighbours can provide it with the highest quality data. Each sensor node maintains and updates its neighbouring table during this phase. Fig.1 shows the structure of the HELLO message.

Source ID	Residual Energy	Free Buffer	Link Performance

Fig. 1. HELLO message structure

C. Link Cost Metric

The link Cost metric is used by the node to select the next hop during the path discovery phase. Let *Ni* be the set of neighbours of node *i*. Then our Cost metric includes an energy factor, available buffer factor and link performance factor that can be computed as below:

$$Cost\ metric = \{ E_{resd,j} + B_{buffer,j} + Lp_{\ ij} \} \qquad (1)$$

Where, $E_{resd,j}$ is the current residual energy of node *j*, where $j \in N_i$, $B_{buffer,j}$ is the available buffer size of node *j* and $Lp_{\ ij}$ is the link performance value between *i* and *j* which is obtained by (2)

$$Lp_{ij} = SNR_{ij} / Distance_{j\ to\ sink} \qquad (2)$$

In here SNR_{ij} is the signal to noise ratio (SNR) for the link between *i* and *j* as well as $Distance_{\text{j to sink}}$ is the distance from node where $j \in Ni$ to sink. So, to select next hop we use from (3).

$$Next\ hop = Max\ \{ Cost\ metric \} \qquad (3)$$

The total Cost metric for a path *P* consists of a set of *K* nodes is the sum of the individual link Cost metrics *l (ij)* along the path. Then the total Cost merit is calculated by (4).

$$CM_{total,p} = \sum_{n=1}^{K-1} l_{(ij)_n} \qquad (4)$$

After initialization phase, each sensor node has enough information to compute the Cost metric for its neighbouring nodes. Then, the sink node locally computes its preferred next hop node using the link Cost metric and sends out a RREQ message to its the most preferred next hop , Fig. 2 shows the structure of the RREQ message . Similarly, through the link Cost metric, the preferred next hop node of the sink computes locally its the most preferred next hop in the direction of the source node and sends out a RREQ message to its next hop, the operation continues until source node.

Source ID	Destination ID	Route ID	Cost Metric	TR	Delay

Fig. 2. RREQ message structure

TR field shows the received time of the packet and Delay field shows the transmission delay of the packet, so we can compute the link end to end delay by using the information in the RREQ message as the source node sends the RREQ message and when an intermediate node N1 receives this RREQ message from the source node, it saves the time of this event in the TR1 field and forwards it to its the most preferred next hop. When a neighbour node (N2) receives the RREQ message from N1, it calculates the difference between the value of TR1 field and the current time (TR2), which represents the measured delay of the link between N1 and N2 as well as stores it in the Delay field.

For the second alternate path, the sink sends alternate path RREQ message to its next the most preferred neighbour. To avoid having paths with shared node, we limit each node to accept only one RREQ message. For those nodes that receive more than one RREQ message only accept the first RREQ message and reject the remaining messages. In order to save energy, we reduce the overhead traffic through reducing control messages. Therefore, instead of periodically flooding a KEEPALIVE message to keep multiple paths alive and update Cost metrics, we append the metrics on the data message by attaching the residual energy, remaining buffer size and link performance to the data message.

D. Paths Selection

After the completion of paths discovery phase, we need to select a set of paths to transfer the traffic from the source to the destination. So out of the *P* paths, the protocol picks out a number of *r* paths to be used to transfer the real-time traffic and *n* paths for non-real-time traffic, where $P = r + n$. To calculate *r*, we assume that the sensor node knows the size of its traffic (both real-time and non-real-time traffic). Let T_r represents the size of the real-time traffic and T_{nr} represents the size of the non-real-time traffic, then we have:

$$r = \frac{T_r}{T_r + T_{nr}} P \qquad (5)$$
$$n = \frac{T_{nr}}{T_r + T_{nr}} P$$

As we divided the *P* paths between the real-time and non-real-time traffic according to the traffic size, we select the best *r* paths that minimize the end to end delay to transfer the real-time traffic to ensure that the critical-time data is delivered to the destination within the time requirements, with out any delay. To find the best baths in terms of the end-to-end delay, during the paths discovery phase, we use Delay field in RREQ message.

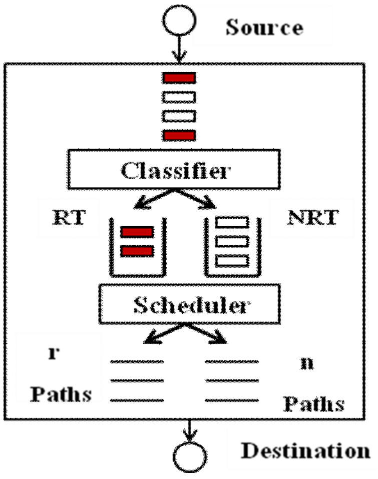


Fig. 3. Functional diagram of the OEQM

E. Traffic Allocation and Data Transmission

OEQM employs the queuing model presented in [18] to handle both real-time and non-real-time traffic. Two different queues are used; one instant priority queue for real-time traffic and the other queue follow the first in first out basis for non-real-time traffic. Fig. 3 shows the functional diagram of the OEQM. The source node knows the degree of the importance of each data packet it is sending which can be translated into predefined priority levels. The application layer sets the required priority level for each data packet by appending an extra bit of information to act as a stamp to distinguish between real-time and non-real-time packets. Based on the packet type, the classifier directs packets into the appropriate queue. The traffic allocation scheme first adds error correction codes to improve the reliability of transmission and to increase the resiliency to paths failures and ensure that an essential portion of the packet is received by the destination without incurring any delay and more energy consumption through data retransmission .Then schedules packets simultaneously for transmission across the available multiple paths . Correction codes are calculated as a function of the information bits to provide redundant information. We use an XOR-based coding algorithm like the one presented in [19]. This algorithm does not require high computation power or high storage space.

After the selection of a set of multiple paths for both traffic types and after adding FEC codes, the source node can begin sending data to the destination along the paths. We use a weighted traffic allocation strategy to distribute the traffic amongst the available paths to improve the end to end delay and throughput. In this strategy, the source node distributes the traffic amongst the paths according to the end to end delay of each path. The end to end delay of each path is obtained during the paths discovery phase via Delay field in RREQ message. Fig. 4 shows the packet format and fields in each segment.

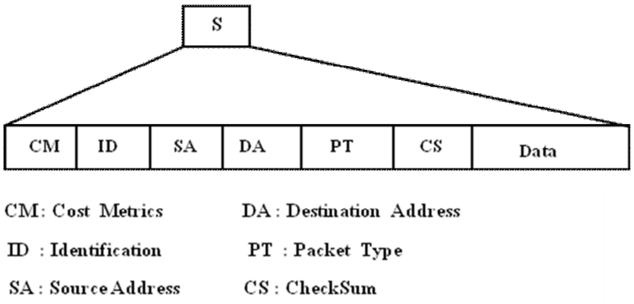


Fig. 4. Packet format

The CM field is an encoded peace of information that represents the current value of metrics used in the Cost metric to avoid excessive control packets to keep routes alive. Each node along the path, after updating its neighbouring table with this information, changes this value by its current metrics.

IV. Performance Evaluation

In this section, we present and discuss the simulation results for the performance evaluation of our protocol. We used NS-2 [20] to implement and simulate OEQM and compare it with the MCMP protocol [14]. Simulation parameters are presented in Table 1 and obtained results are shown below. We investigate the performance of the OEQM in a multi-hop network topology. The metrics used in the evaluation are the energy consumption, delivery ratio and average end to end delay. The average energy consumption is the average of the energy consumed by the nodes participating in message transfer from source node to the sink node. The delivery ratio is the number of packets generated by the source node to the number of packets received by the sink node. The average end to end delay is the average time required to transfer a data packet from source node to the sink node. We study the impact of changing the packet arrival rate on these performance metrics and node failure probability on average energy consumption. Simulation results are averaged over several simulation runs.

A. Impact of packets arrival rate

We change the packet arrival rate at the source node from 5 to 50 packets/sec. The generated traffic at the source node is mixed traffic of both real-time and non-real-time traffic. The real-time traffic is set to 10% of the generated traffic.

TABLE I SIMULATION PARAMETERS

Parameters	Value
Network area	400 m × 400 m
Number of sensors	200
Transmission range	25 m
Packet size	1024 bytes
Transmit power	15 mW
Receive power	13 mW
Idle power	12 mW
Initial battery power	100 J
MAC layer	IEEE 802.11
Max buffer size	256 K-bytes
Simulation time	1000 s

1) Average end to end delay

End to end delay is an important metric in evaluating QoS based routing protocols. The average end to end delay of OEQM and MCMP protocol as the packet arrival rate increases is illustrated in Fig.5. From the results, it is clear that OEQM successfully differentiates network service by giving high real-time traffic absolute preferential treatment over low priority traffic. The real-time traffic is always combined with low end-to-end delay. MCMP protocol outperforms OEQM in the case of non-real-time traffic, because of the overhead caused by the queuing model. Furthermore, for higher traffic rates the average delay increases because the our protocol gives priority to process real-time traffic first, which causes more queuing delay for non-real-time traffic at each sensor node.

2) Packet delivery ratio

Another important metric in evaluating routing protocols is the average delivery ratio. Fig. 6 shows the average delivery ratio of OEQM and MCMP protocols. Obviously, OEQM outperforms the MCMP protocol; this is because in the case of path failures, our protocol uses Forward Error Correction (FEC) technique to retrieve the original message, which is not implemented in the MCMP protocol. Implementing a FEC technique in the routing algorithm enhances the delivery ratio of the protocol as well as minimizes the overall energy consumption especially in the case of route failures.

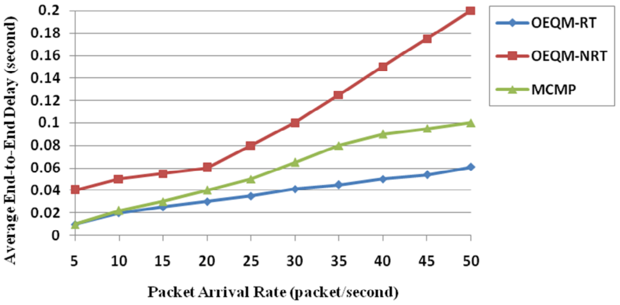


Fig. 5. Average end-to-end delay

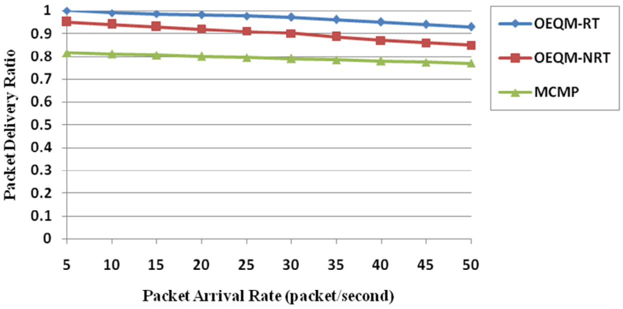


Fig. 6. Packets delivery ratio

3) Average energy consumption

Fig. 7 shows the results for the energy consumption. From the figure, we note that MCMP slightly outperforms OEQM, this is because of the overhead induced by the queuing model and error codes computation. However, meeting the quality of service requirements introduces a certain overhead in terms of energy consumption. Thus minimum tradeoffs with delay and throughput should be made to reduce the energy expenditure. By changing the network conditions and considering node failures, the energy consumption of the MCMP protocol increases significantly as shown in Fig. 8

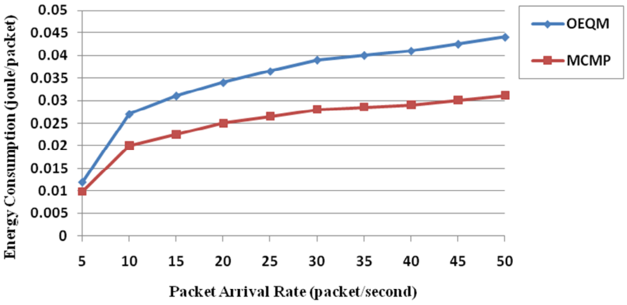


Fig. 7. Average energy consumption

B. Impact of node failure probability

We study the behaviour of protocols in the presence of node failures and change the node failure probability from 0 to 0.05. The results are averaged over several simulation runs. Fig. 8 shows the results for the energy consumption under node failures. Obviously OEQM outperforms the MCMP protocol in this case. Compared to Fig. 7, we observe that OEQM achieves more energy savings than MCMP protocol. This is because our protocol easily recovers from path failures and be able to reconstruct the original messages through the use of the FEC algorithm while the MCMP protocol needs to initiate a data retransmission to recover lost data, which leads to a significant increase in the energy consumption.

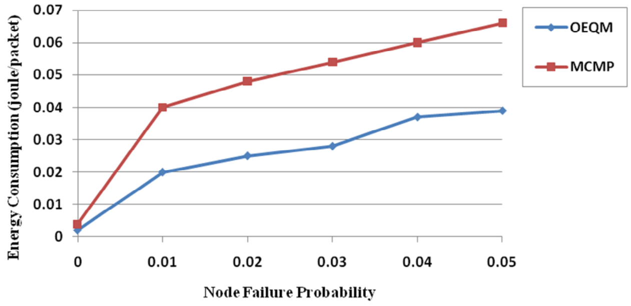


Fig. 8. Average energy consumption as node failure increases

V. Conclusion

In this paper, we have presented an Optimized Energy and QoS Aware Multipath routing protocol (OEQM) in wireless sensor networks to provide service differentiation by giving real-time traffic absolute preferential treatment over the non-real-time traffic. Our protocol uses the multipath paradigm together with a Forward Error Correction (FEC) technique to recover from node failures without invoking network-wide flooding for path-discovery. This feature is very important in sensor networks since flooding consumes energy and consequently reduces the network lifetime.

OEQM uses the residual energy, available buffer size, Signal-to-Noise Ratio (SNR) and distance to sink to predict the best next hop through the paths construction phase also our protocol employs a queuing model to handle both real-time and non-real-time traffic. We have evaluated and studied the performance of our proposed protocol under different network conditions and compared it with the MCMP protocol via NS-2. Simulation results have shown that our protocol achieves lower average delay, more energy savings and higher delivery ratio than the MCMP protocol.

References

[1] K. Srinivasan, M. Ndoh, H. Nie, H. Xia, K. Kaluri and D. Ingraham, “Wireless technologies for condition-based maintenance (CBM) in petroleum plants, ” DCOSS'05, CA, USA, 2005, pp. 389_390.

[2] B. Yahya and J. Ben-Othman, “Towards a classification of energy aware MAC protocols for wireless sensor networks,” Journal of Wireless Communications and Mobile Computing 9 (12), 2009,pp. 1572_1607.

[3] K. Akkaya and M. Younis, “A Survey on Routing for Wireless Sensor Networks”, Journal of Ad Hoc Networks, Vol. 3, 2005, pp. 325- 349.

[4] D. Chen and P.K. Varshney, “QoS Support in Wireless Sensor Networks: a Survey”, In the Proceedings of the International Conference on Wireless Networks (ICWN), 2004, pp. 227-233.

[5] J. N. Al-Karaki and A. E. Kamal, “Routing Techniques in Wireless Sensor networks: A Survey”, IEEE Journal of Wireless Communications, Vol.11, Issue 6, . 2004 , pp. 6 – 28.

[6] A. Martirosyan, A. Boukerche and R. W. N. Pazzi, “A Taxonomy of Cluster-Based Routing Protocols for Wireless Sensor Networks,” ISPAN 2008, pp. 247-253.

[7] A. Martirosyan, A. Boukerche and R. W. N. Pazzi, “ Energy-aware and quality of service-based routing in wireless sensor networks and vehicular ad hoc networks, ” Annales des Telecommunications 63,2008, pp. 669-681.

[8] J. Tsai and T. Moors, “A Review of Multipath Routing Protocols: From Wireless Ad Hoc to Mesh Networks”, 2006.

[9] K. Sohrabi and J. Pottie, “Protocols for self-organization of a wirless sensor network”, IEEE Personal Communications,Vol. 7, Issue 5, 2000, pp 16-27.

[10] K. Akkaya and M. Younis, “An energy aware QoS routing protocol for wireless sensor networks”, In the Proceedings of the MWN, Providence, 2003, pp. 710-715.

[11] M. Younis, M. Youssef and K. Arisha, “Energy aware routing in cluster based sensor networks”, MASCOTS, 2002.

[12] T. He et al., “SPEED: A stateless protocol for real-time communication in sensor networks,” In the Procedings of the Internation Conference on Distributed Computing Systems, Providence, RI, 2003.

[13] E. Felemban, C. G. Lee and E. Ekici, “MMSPEED: multipath multispeed protocol for QoS guarantee of reliability and timeliness in wireless sensor networks,” IEEE Trans. on Mobile Computing, vol. 5, no. 6, 2006, pp. 738–754.

[14] X. Huang and Y. Fang, “Multiconstrained QoS Mutlipath Routing in Wireless Sensor Networks,” Wireless Networks ,2008, 14:465-478.

[15] A. B. Bagula and K. G. Mazandu,”Energy Constrained Multipath Routing in Wireless Sensor Networks”, UIC 2008, LNCS 5061, pp 453-467.

[16] D. Ganesan, R. Govindan, S. Shenker and D. Estrin, “Energy-efficient multipath routing in wireless sensor networks, ” ACM SIGMOBILE Mobile Computing and Communications Review 5 (4) ,2001,pp. 11_25.

[17] Ch. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann and F. Silva, “Directed diffusion for wireless sensor networking, ” ACM/IEEE Transactions on Networking (TON) 11 (1) ,2002, pp. 2_16.

[18] K. Akkaya and M. Younis, “An energy aware QoS routing protocol for wireless sensor networks, ” In the Proceedings of the 23rd International Conference on Distributed Computing Systems Workshops, Providence, RI, USA, 2003, pp. 710_715.

[19] Z. Xiong, Z. Yang, W. Liu and Z. Feng, “A lightweight FEC algorithm for fault tolerant routing in wireless sensor networks, ” WiCOM-2006, 2006, pp. 1-4.

[20] VINT, The network simulator – ns-2, 2008, <http://www.isi.edu/nsnam/ns/>.

A Hybrid Approach for DICOM Image Feature Extraction, Feature Selection Using Fuzzy Rough set and Genetic Algorithm

J. Umamaheswari

Research Scholar, Department of Computer Science
Dr. G.R.D College of Science,
Coimbatore, Tamilnadu, India
Umamugesh@yahoo.com

Dr. G. Radhamani

Director, Department of Computer Science
Dr. G.R.D College of Science,
Coimbatore, Tamilnadu, India.
radhamanig@gmail.com

Abstract— **The proposed hybrid approach for feature extraction, feature reduction and feature selection of Medical images based on Rough set and Genetic Algorithm (GA). A Gray Level Co-occurrence Matrix (GLCM) and Histogram based texture feature set is derived. The optimal texture features are extracted from normal and infected Digital Imaging and Communications in Medicine (DICOM) images by using GLCM and histogram based features. The inputs of these features are taken for the feature selection process. The selected features is solved by using Fuzzy Rough set and GA. These optimal features are used to classify the DICOM images into normal and infected. The performance of the algorithm is evaluated on a series of DICOM datasets collected from medical laboratories.**

Keywords- Fuzzy roughest; GLCM; Texture features; Histogram Features and region features.

I. INTRODUCTION

Nowadays DICOM image analysis is becoming more important for diagnosis process. This process is not easy way for optimal identification and early detection of diseases for improving the surviving rate. Generally the DICOM image is a valuable and most reliable method in early detection.

Different methods of DICOM image feature reduction have been used to solve by statistical methods, texture based methods and feature is extracted by using image processing techniques [3]. Some other methods are based on fuzzy theory [1] and neural networks [2].

The lack of systematic research on features extracted and their role to the classification results forces researchers to select features arbitrarily as input to their systems. Genetic algorithms have been successful in discovering an optimal or near-optimal solution amongst a huge number of possible solutions (Goldberg 1989). Moreover, a combination of genetic algorithms and fuzzy can prove to be very powerful in classification problems. Previously genetic algorithms have been used either to evolve neural network topology (Stathakis and Kanellopoulos 2006) [4] or to select features (Kavzoglou and Mather 2002 [5]) but not both at the same time.

GLCM, Histogram, level set, Gabor filters, and wavelet transform [6, 7, 8, 9] are the approaches for texture classification problem. The Gabor filters are poor due to their lack of orthogonality that results in redundant features, while wavelet transform is capable of representing textures at the most suitable scale, by varying the spatial resolution and there is also a wide range of choices for the wavelet function.

In medical image analysis, the determination of normal and infected brain is classified by using texture. DICOM and CT image texture proved to be useful to determine the Normal brain [10] and to detect the brain disease part [11].

There is a big problem in selecting the optimal features in medical imaging. The evaluation of possible feature subsets is usually a painful task. So the large amount of computational effort is required. Fuzzy roughest and Genetic algorithm (GA) appear to be a selective approach to choose the best feature subset while maintaining acceptable feature selection. Siedlecki and Sklansky [12] compared the GA with classical algorithms and they proposed the GA for feature selection. Fuzzy rough set proved to be the best selection method for optimal classification.

A new method for extracting features in DICOM images with lower computational requirements is proposed and selection percentage is analyzed. The tables provide the user with all relevant information for taking efficient decision. Thus a synergy of genetic algorithms and fuzzy is used for feature selection in our proposed method.

The remaining paper is organized as follows. Section 2 describes the feature extraction process. The feature selection problem is discussed in Section 3, while Section 4 contains the experimental results. Finally section 5 presents conclusion and references.

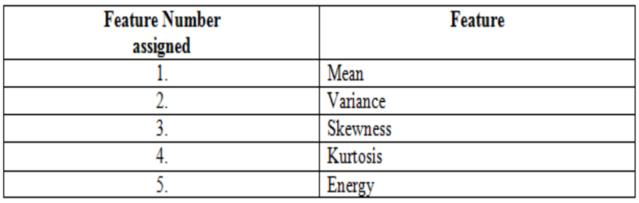
II. Feature Extraction

Feature extraction methodologies analyze objects and images to extract the features that are representative of the various classes of objects. In this Work intensity histogram features and Gray Level Co-Occurrence Matrix (GLCM) features are extracted [12].

2.1 Intensity Histogram Features

Intensity Histogram analysis has been extensively used. The intensity histogram features are mean, variance, skewness, kurtosis, entropy and energy. These are shown in Table 1.

TABLE 1 Features of Intensity Histogram



Feature Number assigned	Feature
1.	Mean
2.	Variance
3.	Skewness
4.	Kurtosis
5.	Energy

The average value of intensity histogram features obtained for different type of medical image is given Table 2 as follows:

TABLE 2 Intensity Histogram Features for Medical Images

	Mean	Variance	Skewness	kurtosis	Energy
Normal	7.976583	0.09123	18.4427	380.7879	0.986126
Infected	7.680339	0.059461	-53.2573	649.3231	0.991122

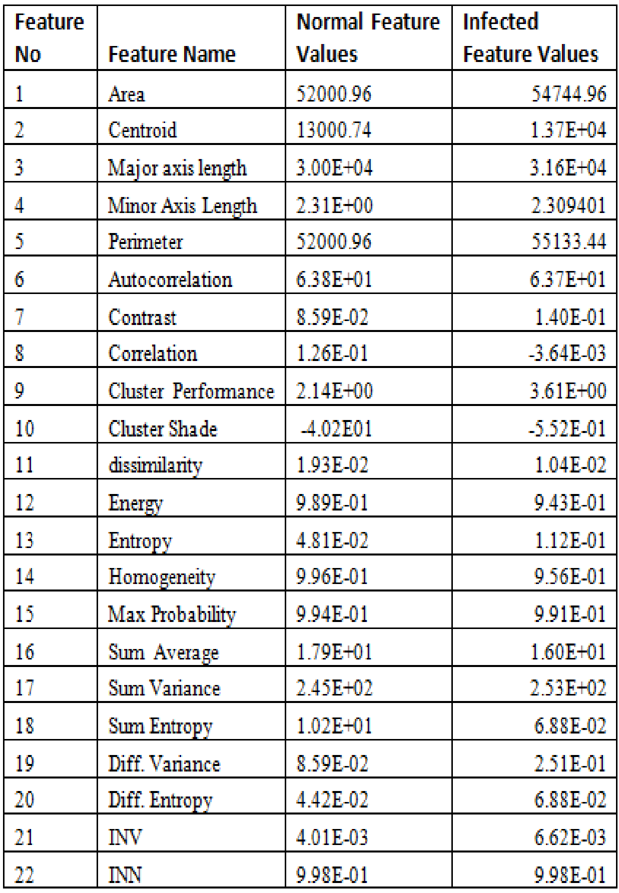
2.2 GLCM Features

The Gray-Level Co-occurrence Matrix (GLCM) is a statistical method that considers the spatial relationship of pixels, which is also known as the gray-level spatial dependence matrix. The pixel and the adjacent pixel is consider as the spatial relationship and also another spatial relationships can be specified between these two pixels.

The Following GLCM features were extracted in this paper : Autocorrelation, Contrast, Correlation, Cluster Prominence, Cluster Shade, Dissimilarity Energy, Entropy, Homogeneity, Maximum probability, Sum of squares, Sum average, Sum variance, Sum entropy, Difference variance, Difference entropy, Information measure of correlation, information measure of correlation, Inverse difference normalized.

The value obtained for the above features for a typical normal and infected DICOM image is given in the following Table 3,

TABLE 3 GLCM Features and Values Extracted From Normal & Infected Medical Images



Feature No	Feature Name	Normal Feature Values	Infected Feature Values
1	Area	52000.96	54744.96
2	Centroid	13000.74	1.37E+04
3	Major axis length	3.00E+04	3.16E+04
4	Minor Axis Length	2.31E+00	2.309401
5	Perimeter	52000.96	55133.44
6	Autocorrelation	6.38E+01	6.37E+01
7	Contrast	8.59E-02	1.40E-01
8	Correlation	1.26E-01	-3.64E-03
9	Cluster Performance	2.14E+00	3.61E+00
10	Cluster Shade	-4.02E01	-5.52E-01
11	dissimilarity	1.93E-02	1.04E-02
12	Energy	9.89E-01	9.43E-01
13	Entropy	4.81E-02	1.12E-01
14	Homogeneity	9.96E-01	9.56E-01
15	Max Probability	9.94E-01	9.91E-01
16	Sum Average	1.79E+01	1.60E+01
17	Sum Variance	2.45E+02	2.53E+02
18	Sum Entropy	1.02E+01	6.88E-02
19	Diff. Variance	8.59E-02	2.51E-01
20	Diff. Entropy	4.42E-02	6.88E-02
21	INV	4.01E-03	6.62E-03
22	INN	9.98E-01	9.98E-01

III. Feature Selection

To improve the prediction accuracy and minimize the computation time, feature selection is used. Feature selection occurs by reducing the feature space. This is achieved by removing irrelevant, redundant and noisy features which performs the dimensionality reduction. Popularly used feature selection algorithms are Sequential forward Selection, Sequential Backward selection, Genetic Algorithm and Particle Swarm Optimization. In this paper a combined approach of fuzzy roughest method with Genetic Algorithm is proposed to select the optimal features. The selected optimal features are considered for classification.

3.1 Genetic Algorithm (GA) based Feature selection:

During classification, the number of features can be large, irrelevant or redundant. So the optimal solution is not occurred. To solve this problem feature reduction is introduced to improve the process by searching for the best features subset, from the original features.

GA is an adaptive method of global-optimization searching and simulates the behavior of the evolution process in nature. It is based on Darwin's fittest principle, which states that an initial population of individuals evolves through natural selection in such a way that the fittest individuals have a higher chance of survival.

The GA maintains a cluster of competing feature matrices. To evaluate each matrix in this cluster, the inputs are multiplied by the matrix, producing a set of output which are then sent to a classifier. The classifier typically divides the features into a training set and a testing set, to evaluate classification accuracy. Generally each feature is encoded into a vector called a chromosome.

$$\text{fitness} = W_A \cdot \text{Accuracy} + W_{nb}/N \qquad (1)$$

where W_A is the weight of accuracy and W_{nb} is the weight of N feature participated in classification where $N \neq 0$.

A fitness value will be used to measure the fitness of a chromosome and decides whether a chromosome is good or not in a given cluster. Initial populations in the genetic process are randomly created. GA uses three operators to produce a next generation from the current generation: ***reproduction***, ***crossover*** and ***mutation***. GA eliminates the chromosomes of low fitness and keeps the ones of high fitness.

Thus more chromosomes of high fitness move to the next generation. This process is repeated until a good chromosome (individual) is found. The Figure 1 illustrates the feature selection using the genetic algorithm.

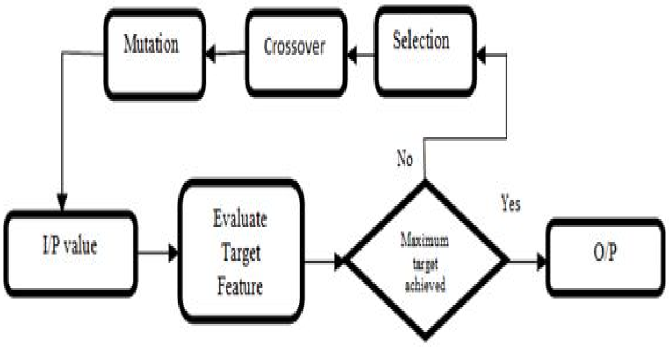
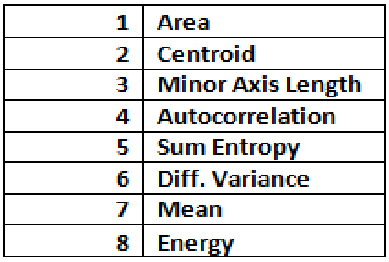


FIGURE 1 FEATURE SELECTION USING GA

The total features extracted are 40. The selected features using GA method are tabulated as follows:

TABLE 4 FEATURE SELECTED BY GENETIC ALGORITHM METHOD



1	Area
2	Centroid
3	Minor Axis Length
4	Autocorrelation
5	Sum Entropy
6	Diff. Variance
7	Mean
8	Energy

The above Table 5 shows the feature selected by GA method.

3.2 Feature selection by Rough Set

Fuzzy set involves more advanced mathematical concepts, real numbers and functions, whereas in classical set theory the notion of a set is used as a fundamental notion of whole mathematics and is used to derive any other mathematical concepts, e.g., numbers and functions [13,14].

Rough set theory can be viewed as a specific implementation of Frege's idea of vagueness, i.e., imprecision in this approach is expressed by a boundary region of a set, and not by a partial membership, like in fuzzy set theory. Rough set concept can be defined quite generally by means of topological operations, interior and closure, called approximations. The concept of rough set theory is based on the followings:

3.2.1 Decision Tables

A decision table consists of two different attribute sets. One attribute set is designated to represent Conditions (C) and another set is to represent Decision (D). Therefore, each row of a decision table describes a decision rule, which indicates a particular decision to be taken if its corresponding condition is satisfied. If a set of decision rules has common condition but different decisions then all the decision rules belonging to this set are inconsistent decisions, otherwise; they are consistent.

3.2.2 Dependency of Attributes

Similar to relational databases, dependencies between attributes may be discovered. If all the values of attributes from D are uniquely determined by values of attributes from C then D depends totally on C or C functionally determines D which is denoted by $C \Rightarrow D$. If D depends on some of the attributes of C (i.e. not on all) then it is a partial dependency $C \Rightarrow_k D$ and a degree of dependency (k; $0 \leq k \leq 1$) can be computed as $k = \gamma(C, D)$, where $\gamma(C, D)$ is the consistency factor of the decision table. $\gamma(C, D)$ is defined as the ratio of the number of consistent decision rules to the total number of decision rules in the decision tables.

3.2.3 Reduction of Attributes

Decision tables where feature vectors are the condition (C) and desired values for corresponding classes are the decisions (D) can also represent classification of feature vectors. Now the dimensionality reduction can simply be considered as removal of some attributes from the decision table (actually some features from the feature vector) preserving its basic classification capability. If a decision table contains some redundant or superfluous data, then collect those redundant data and remove them.

The selected features using Rough set method are tabulated as follows

TABLE 5 FEATURE SELECTED BY ROUGH SET METHOD

1	**Kurtosis**
2	**Std**
3	**Sum Average**
4	**Sum Variance**

3.3 Proposed Hybrid Approach Algorithm:

1. N number of features is extracted by GLCM and Histogram texture features from the preprocessed Image
2. Apply roughest algorithm to select the optimal set containing n1 number of features where n1< N
3. Apply genetic algorithm to select the best subset containing n2 number of features where n2<N
4. Find the Union of n1 features and n2 features to form final n features
5. Use the n features where n<N for Classification.

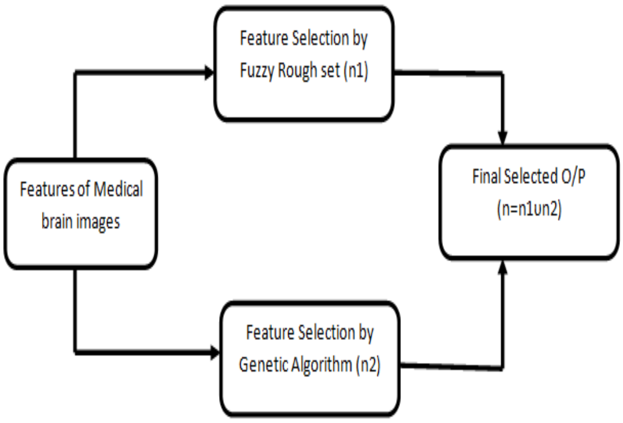
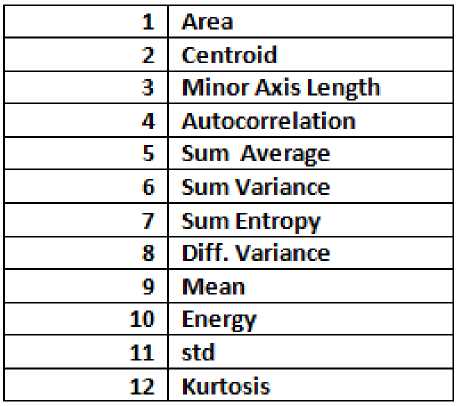

Feature Selection by Fuzzy Rough set (n1)
Features of Medical brain images
Final Selected O/P (n=n1un2)
Feature Selection by Genetic Algorithm (n2)

FIGURE 2 PROPOSED APPROACH FOR FEATURE SELECTION

The above Figure 2 shows the feature selection by proposed approach. The following Table 6 gives feature selected by proposed approach.

TABLE 6 FEATURE SELECTED BY PROPOSED APPROACH



1	**Area**
2	**Centroid**
3	**Minor Axis Length**
4	**Autocorrelation**
5	**Sum Average**
6	**Sum Variance**
7	**Sum Entropy**
8	**Diff. Variance**
9	**Mean**
10	**Energy**
11	**std**
12	**Kurtosis**

IV. EXPERIMENTAL RESULTS

For the comparison of results of different feature reduction methods like rough set, GA and the proposed method has been used. Feature space is formed using the DICOM images. Totally forty features are extracted which forms the feature space. Using GA feature space reduced to eight features and by rough set method it is reduced to four features. The proposed method selects only twelve features. These features improve the class prediction.

The percentage of reduction by GA method is 80%. 75 % of reduction is done by rough set method. The selected features are used for classification which reduces the classification time and improves the prediction accuracy. The proposed approach selects feature space of DICOM images which is reduced by 95%. The following Table 7 gives the results of the proposed method.

TABLE 7 RESULTS OBTAINED BY PROPOSED METHOD

GA method	80%
Rough set Method	75%
Proposed method	95%

This gives that the proposed approach is efficient for image analysis. It's a better tool for doctors or radiologists to classify normal brain images and infected brain images.

V. CONCLUSION

The paper developed a hybrid technique with normal and infected DICOM images. The proposed approach gives results in extraction and selection for classifying the images that benefit the physician to make a final decision. The approach for feature extraction, feature reduction and feature selection of images based on Rough set and Genetic Algorithm (GA). A Gray Level Co-occurrence Matrix (GLCM) and Histogram based texture feature set is derived. The feature selection is done by Fuzzy Rough set and GA. These optimal features are used to classify the DICOM images into normal and infected.

The performance of the algorithm is evaluated on a series of DICOM datasets collected from medical laboratories. The method has been proved that it is easier and gives desirable results for future process.

REFERENCES

[1] D.Brazokovic and M.Nescovic ., "Mammogram screening using multisolution based image segmentation", International journal of pattern recognition and Artificial Intelligence, Vol.7,No.6, P. 1437-1460,1993.

[2] I.Christiyanni et al ., "Fast detection of masses in computer aided mammography", IEEE Signal processing Magazine, P.54- 64, 2000.

[3] S.Lai,X.Li and W.Bischof . "On techniques for detecting circumscribed masses in mammograms", IEEE Trans on Medical Imaging, Vol.8, No.4, P.377-386,1989.

[4] K. Topouzelis, D. Stathakis **and** V. Karathanassi , **"**Investigation of genetic algorithms contribution to feature selection for oil spill detection**"**, **Vol.** 30, **No.**3, P.611-625, 2009**.**

[5] Kavzoglu T and Mather P.M., "The role of feature selection in artificial neural network applications", International Journal of Remote Sensing, Vol.23, No.15, P.2919-2937, 2002.

[6] Dunn C., Higgins W.E., "Optimal Gabor filters for texture segmentation", IEEE Transactions on Image Processing, Vol. 4, No.7, P. 947-964,1995.

[7] Chang T., Kuo C., "Texture Analysis and classification with tree structured wavelet transform", IEEE Transactions on Image Processing, Vol. 2, No.4, P. 429-441, 1993.

[8] Dr. H.B.Kekre, Sudeep D. Thepade, Tanuja K. Sarode and Vashali Suryawanshi, " Image Retrieval using Texture Features extracted from GLCM, LBG and KPE", Vol. 2, No. 5, P.1793-8201, 2010.

[9] M.M. Trivedi, R.M. Haralick, R.W. Conners, and S. Goh, "Object Detection based on Gray Level Coocurrence", Computer Vision, Graphics, and Image Processing, Vol. 28, P. 199-219, 1984.

[10] Schad L.R., Bluml S., Zuna, I., "MR tissue characterization of intracranial tumors by means of texture analysis, Magnetic Resonance Imaging", Vol.11, No.6, P. 889-896, 1993.

[11] Free borough P.A., Fox N.C., "MR image texture analysis applied to the diagnosis and tracking of Alzheimer's disease ", IEEE Transactions on Medical Imaging, Vol. 17, No.3, P. 475-479, 1998.

[12] Serkawt Khola , "Feature Weighting and Selection A Novel Genetic Evolutionary Approach", World Academy of Science, Engineering and Technology 73, P.1007-1012, 2011.

[13] Ping Yao, "Fuzzy Rough Set and Information Entropy Based Feature Selection for Credit Scoring", IEEE , P.247-251, 2009.

[14] Pradipta Maji and Sankar K. Pal, "Fuzzy–Rough Sets for Information Measures andSelection of Relevant Genes From Microarray Data", IEEE, Vol. 40, No. 3, P.741-752, 2010.

AUTHORS PROFILE

Ms.J.Umamaheswari, Research Scholar in Computer Science, Dr. GRD college, Coimbatore. She has 5 years of teaching experience and two years in Research. Her areas of interest include Image Processing, Multimedia and communication. She has more than 3 publications at International level. She is a life member of professional organization IAENG.

Dr.G. Radhamani, Director in Computer Science, Dr. GRD College, Coimbatore. She has more than 5 years of teaching and research experience. She has volume of publications at International level. Her areas of interest include Mobile computing, e-internet and communication. She is a member of IEEE.

Studying the Performance of Transmitting Video Streaming over Computer Networks in Real Time

Hassan H. Soliman

Department of Electronics and
Communication Engineering,
Faculty of Engineering,
Mansoura University, EGYPT

Hazem M. El-Bakry

Department of Information Systems,
Faculty of Computer Science &
Information Systems, Mansoura
University, EGYPT
helbakry20@yahoo.com

Mona Reda

Senior multimedia designer, E-learning unit, Mansoura University,
EGYPT

Abstract—**the growth of Internet applications has become widely used in many different fields. Such growth has motivated video communication over best-effort packet networks. Multimedia communications have emerged as a major research and development area. In particular, computers in multimedia open a wide range of possibilities by combining different types of digital media such as text, graphics, audio, and video. This paper concentrates on the transmission of video streaming over computer networks. This study is preformed on two different codecs H.264 and MPEG-2. Video streaming files are transmitted by using two different protocols HTTP and UDP. After making the real time implementation, the performance of transmission parameters over the computer network is measured. Practical results show that jitter time of MPEG-2 is less than H.264. So MPEG-2 protocol is better than H.264 over the UDP protocols. In contrast, jitter time of H.264 is less than MPEG-2 over HTTP protocol. So H.264 is better than MPEG-2 over the HTTP protocol. This is from the network performance view. However, from video quality view, MPEG-2 achieves the guidelines of QoS of video streaming.**

Keywords- Multimedia communication, Video streaming, Network performance

I. INTRODUCTION

Multimedia is one of the most important aspects of the information era. It can be defined as a computer based interactive communications process that incorporates text, graphics, animation, video and audio. Due to the rapid growth of multimedia communication, multimedia standards have received much attention during the last decade. Multimedia communications have been emerged as a major research and development area. In particular, computers in multimedia open a wide range of possibilities by combining different types of digital media such as text, graphics, audio, and video.

The growth of the Internet in the mid-1990's motivated video communication over best-effort packet networks. Multimedia provides an environment in which the user can interact with the program.

There are two different playout methods allow covering of the (Audio/Video) A/V streaming requirements.

1. Streaming from File: Audio and video are encoded and stored in a file. The file is then scheduled for later broadcast and uploaded to the operator of the distribution network. At the scheduled broadcast time, the playout begins from the media file stored at the broadcaster's location. This scheduling method is particularly useful, when a media event has been prerecorded some time before the broadcast is scheduled.

2. Live Event Streaming: is, as the name says, a vehicle for broadcasting streams covering live events. The broadcast is scheduled exactly as in the file propagation method. A video camera at the location of the event captures the event, and an encoder converts the video stream into an MPEG stream. At the time of the broadcast, this stream is accepted on a TCP/IP port at the broadcaster's location (assuming that the system is IP based). The stream is then wrapped into subscription packages and replicated onto the broadcast stream. The advantage of this is that the content is not stored anywhere and is directly broadcast [1].

The motivation of this paper is to send video streaming over the network, and find the suitable protocol and also best codec in transmission.

The paper organization as the following: section related work is consider as a short description about the codecs types. Section video streaming implementation gives a description of platform and what is the measurement used in this implementation and display result figures. Section experimental results is summery the result and choose the best codec used over the suitable transmission protocols. Finally, the conclusion of this paper.

II. RELATED WORK

Noriaki Kamiyama [2] is proposed to stream high definition video over the internet. However, the transmission bit-rate is quite large, so generated traffic flows will cause link congestion. Therefore, when providing streaming services of rich content such as videos with HDTV or UHDV quality, it is important to reduce the maximum link utilization. Tarek R. Sheltami [3] is presented a simulation to analysis the performance of wireless networks under video traffic by minimization power and other QoS requirements such as delay jitter. Yung-Sung Huang [4] is proposed video streaming from both video servers in hospitals and webcams localized to patients. All important medical data are transmitted over a 3G-wireless communication system to various client devices. Also, proposed a congestion control scheme for streaming process to reduce packet losses.

This paper concentrates on the transmission of video streaming over computer networks. This study is preformed on two different codecs H.264 and MPEG-2. Video streaming files are transmitted by using two different protocols HTTP and UDP. After making the real time implementation, the performance of transmission parameters over the computer network is measured. Practical results show that jitter time of MPEG-2 is less than H.264. So MPEG-2 protocol is better than H.264 over the UDP protocols. In contrast, jitter time of H.264 is less than MPEG-2 over HTTP protocol. So H.264 is better than MPEG-2 over the HTTP protocol. This is from the network performance view. However, from video quality view, MPEG-2 achieves the guidelines of QoS of video streaming.

III. RELATED VIDEO FORMAT

There are two standards bodies which are responsible to put the video coding standards, the International Standards Organization (ISO) and the International Telecommunications Union (ITU), have developed a series of standards that have shaped the development of the visual communications industry. The ISO JPEG, MPEG-1, MPEG-2, and MPEG-4 standards have perhaps had the biggest impact: JPEG has become one of the most widely used formats for still image storage and *MPEG-2* forms the heart of digital television and DVD-video systems [5].

The ITU's H.261 standard was originally developed for video conferencing over the ISDN, but H.261 and H.263 are now widely used for real-lime video communications over a range of networks including the Internet. H.264 of ITU-T is known as International Standard video coding, it is the latest standard in a sequence of the video coding standards H.261 [6].

Each of the international standards takes a similar approach to meeting these goals. A video coding standard describes syntax for representing compressed video data and the procedure for decoding this data as well as (possibly) a 'reference' decoder and methods of proving conformance with the standard [1].

MPEG-1

The first standard produced by the Moving Picture Experts Group, popularly known as MPEG-1, was designed to provide video and audio compression for storage and playback on CD-ROMs. MPEG-1 aims to compress video and audio to a bit rate 1.4 Mbps with a quality that is comparable to VHS (Video home system) video tape.

MPEG-1 is important for two reasons:

1. It gained widespread use in other video storage and transmission applications (including *CD* storage as part of interactive applications and video playback over the Internet)

2. Its functionality is used and extended in the popular MPEG-2 standard.

The MPEG-1 standard consists of three parts: Part 1: deals with system issues (including the multiplexing of coded video and audio). Part 2: deals with compressed video, video was developed with aim of supporting efficient coding of video for *CD* playback applications and achieving video quality comparable to, or better than, *VHS* videotape at *CD* bit rates (around 1.2Mbps for video). Part 3: deal with compressed audio.

MPEG-2

The next important entertainment application for coded video (after CD-ROM storage) was digital television. It has to efficiently support larger frame sizes (typically 720 x 576 or 720 x 480 pixels for ITU-R 601 resolution) and coding of interlaced video [5].

The MPEG-2 standard was designed to provide the capability for compressing , coding, and transmitting high-quality, multichannel multimedia signals over terrestrial broadcast, satellite distribution, and broadband networks [7]. MPEG-2 consists of three main sections: video, audio (based on MPEG-1 audio coding) and, systems (defining, in more detail. than MPEG-1 systems, multiplexing and transmission of the coded audio/visual stream).

MPEG-2 video is a superset of MPEG-1 video; most MPEG-1 video sequences should be decodable by an MPEG-2 decoder. There are 4 main enhancements added by the MPEG-2 standard are as follows: Efficient coding of television-quality video, support for coding of interlaced video, scalability, profiles and levels.

Efficient coding of television-quality video: The most important application of MPEG-2 is broadcast digital television. The ‘Core’ functions of MPEG-2 are optimized for efficient coding of television resolutions at a bit rate of around 3-5 Mbps.

Support for coding of interlaced video: MPEG-2 video has several features that support flexible coding of interlaced video. The two fields that make up a complete interlaced frame can be encoded as separate pictures (field pictures), each of which is coded as an I-, P- or B-picture. P- and B-field pictures may be predicted from a field in another frame or from the other field in the current frame.

Alternatively, the two fields may be handled as a single picture (a frame picture) with the luminance samples in each macroblock of a frame picture arranged in one of two ways as figure1. Frame DCT coding is similar to the MPEG-1 structure, where each of the four luminance blocks contains alternate lines from both fields. With field DCT coding, the top two luminance blocks contain only samples from the top field, and the bottom two luminance blocks contain samples from the bottom field.

In a field picture, the upper and lower 16 x 8 sample regions of a macroblock may be motion-compensated independently: hence each of the two regions has its own vector (or two vectors in the case of a B-picture). However, this 16 x 8 motion compensation mode can improve performance because a field picture has half the vertical resolution of a frame picture and so there are more likely to be significant differences in motion between the top and bottom halves of each macroblock.

Scalability: A scalable coded bit stream consists of a number of layers, a base layer and one or more enhancement layers. The base layer can be decoded to provide a recognizable video sequence that has a limited visual quality, and a higher-quality sequence may be produced try decoding the base layer plus enhancement layer(s), with each extra enhancement layer improving the quality of the decoded sequence. MPEG-2 video supports 4 scalable modes: spatial scalability, temporal scalability, SNR scalability, and data partitioning [5].

Profiles and levels: With MPEG 2, *profiles* specify the syntax (i.e., algorithms) and *levels* specify various parameters (resolution, frame rate, bit rate, etc.).

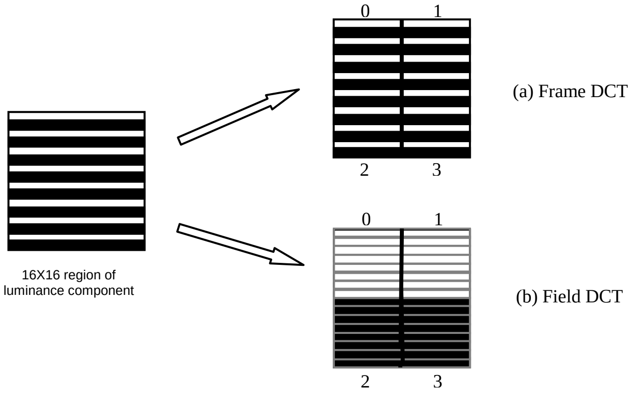


Figure1. Illustration of the two coding structures.

Levels: MPEG 2 supports four levels, which specify resolution, frame rate, coded bit rate, and so on for a given profile. 1. Low Level (LL) MPEG 1 Constrained Parameters Bitstream (CPB) supports up to 352 × 288 at up to 30 frames per second. Maximum bit rate is 4 Mbps.

Main Level (ML): MPEG 2 Constrained Parameters Bitstream (CPB) supports up to 720 × 576 at up to 30 frames per second and is intended for SDTV applications. Maximum bit rate is 15–20 Mbps.

High 1440 Level: This Level supports up to 1440 × 1088 at up to 60 frames per second and is intended for HDTV applications. Maximum bit rate is 60–80 Mbps.

High Level (HL): High Level supports up to 1920 × 1088 at up to 60 frames per second and is intended for HDTV applications. Maximum bit rate is 80–100 Mbps.

Profiles: MPEG 2 supports six profiles, which specify which coding syntax (algorithms) is used.

1. Simple Profile (SP): Main profile without the B frames, intended for software applications and perhaps digital cable TV.

2. Main Profile (MP): Supported by most MPEG 2 decoder chips, it should satisfy 90% of the SDTV applications. Typical resolutions are shown in Table I [6].

3. Multiview Profile (MVP): By using existing MPEG 2 tools, it is possible to encode video from two cameras shooting the same scene with a small angle between them.

4. 4:2:2 Profile (422P): Previously known as “studio profile,” this profile uses 4:2:2 YCbCr instead of 4:2:0, and with main level, increases the maximum bit rate up to 50 Mbps (300 Mbps with high level). It was added to support pro-video SDTV and HDTV requirements.

5. SNR and Spatial Profiles: Adds support for SNR scalability and/or spatial scalability.

6. High Profile (HP): Supported by MPEG 2 decoder chips targeted for HDTV applications. Typical resolutions are shown in Table I[7].

H.261

ITU-T H.261 was the first video compression and decompression standard developed for video conferencing. The video encoder provides a self-contained digital video bitstream which is multiplexed with other signals, such as control and audio. The video decoder performs the reverse process.

H.261 video data uses the 4:2:0 YCbCr format shown previous, with the primary specifications listed in Table II. The maximum picture rate may be restricted by having 0, 1, 2, or 3 non-transmitted pictures between transmitted ones. Two picture (or frame) types are supported: Intra or I Frame: A frame having no reference frame for prediction. Inter or P Frame: A frame based on a previous frame [5].

H.264

The new video coding standard Recommendation H.264 of ITU-T also known as International Standard 14496-10 or MPEG-4 part 10 Advanced Video Coding (AVC) of ISO/IEC. H.264/AVC was finalized in March 2003 and approved by the ITU-T in May 2003 [3].

H.264/AVC offers a significant improvement on coding efficiency compared to other compression standards such as MPEG-2. The functional blocks of H.264/AVC encoder and decoder are shown in Fig.2 and Fig.3 respectively.

In Fig.2 the sender might choose to preprocess the video using format conversion or enhancement techniques. Then the encoder encodes the video and represents the video as a bit stream.

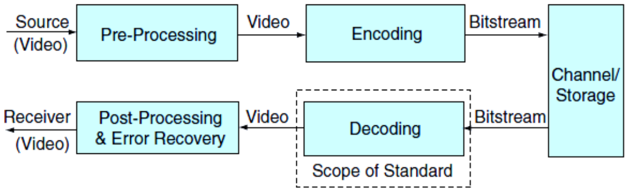


Figure 2. Scope of video coding standardization: Only the syntax and semantics of the bitstream and its decoding are defined.

After transmission of the bit stream over a communications network, the decoder decodes the video which gets displayed after an optional post-processing step which might include format conversion, filtering to suppress coding artifacts, error concealment, or video enhancement.
The standard defines the syntax and semantics of the bit stream as well as the processing that the decoder needs to perform when decoding the bit stream into video, not define how encoding or other video pre-processing is performed thus enabling manufactures to compete with their encoders in areas like cost, coding efficiency, error resilience and error recovery, or hardware requirements.
At the same time, the standardization of the bit stream and the decoder preserves the fundamental requirement for any communications standard—interoperability.

Table I. Example Levels and Resolutions for MPEG 2 Main Profile [6].

Level	Maximum Bit Rate (Mbps)	Typical Active Resolutions	Refresh Rate2 (HZ)	Typical Active Resolution	Refresh Rate2 (HZ)
High	80 (300 for 4:2:2 profile)	1920*10801	23.976p		
			24p		
			25p		
			29.97p		
			30p		
			50i		
			59.94i		
			60i		
High 1440	60	1440*720	23.976p		
			24p		
			25p		
			29.97p		
			30p		
			50p		
			59.94p		
			60p		
		1920*10801	50i		
			50i59.94i		
			59.9i		
			60i		
Main	15 (50 for 4:2:2 profile)		29.97p	352*576	25p
			29.97p	544*576	25p
			29.97p		
			29.97p	704*576	25p
			29.97p	720*576	25p
Low	4	320*240	29.97p		
		352*240	29.97p	352*288	25p

Notes: 1. The video coding system requires that the number of active scan lines be a multiple of 32 for interlaced pictures, and a multiple of 16 for progressive pictures. Thus, for the 1080-line interlaced format, the video encoder and decoder must actually use 1088 lines. The extra eight lines are "dummy" lines having no content, and designers choose dummy data that simplifies the implementation. The extra eight lines are always the last eight lines of the encoded image. These dummy lines do not carry useful information, but add little to the data required for transmission.
2. p = progressive; i = interlaced.

H.264/AVC consists of two conceptual layers (Fig.3). The video coding layer (*VCL*) defines the efficient representation of the video, and the network adaptation layer (NAL) converts the VCL representation into a format suitable for specific transport layers or storage media. For circuit-switched transport like H.320, H.324M or MPEG-2, the NAL delivers the coded video as an ordered stream of bytes containing start codes such that these transport layers and the decoder can robustly and simply identify the structure of the bit stream. For packet switched networks like RTP/IP or TCP/IP, the NAL delivers the coded video in packets without, these start codes.

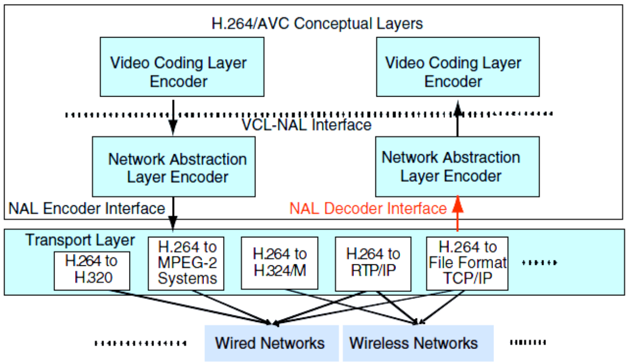


Figure 3. H.264/AVC in a transport environment: The network abstraction layer interface enables a seamless integration with stream and packet-oriented transport layers

H.264/AVC introduces the following changes:
1. In order to reduce the block-artifacts an adaptive deblocking filter is used in the prediction loop. The deblocked macroblock is stored in the memory and can be used to predict future macroblocks.
2. Whereas the memory contains one video frame in previous standards, H.264/AVC allows storing multiple video frames in the memory.
3. In H.264/AVC a prediction scheme is used also in Intra mode that uses the image signal of already transmitted macroblocks of the same image in order to predict the block to code.
4. The Discrete Cosine Transform (DCT) used in former standards is replaced by an integer transform [7].

Table II. H.261 YCbCr Parameters [7].

Parameters	CIF	QCIF
Active resolution (Y)	352 × 288	176× 144
Frame refresh rate	29.97 Hz	
YCbCr sampling structure	4:2:0	
form of YCbCr coding	Uniformly quantized PCM, 8 bits per sample.	

IV. VIDEO STREAMING IMPLEMENTATION

In this paper discussed the implementation platform. There are two computers connected each other via switch as in fig.4. The link connected between computers and switch is 100 Mbps. One PC is used as video streamer and the other one is used as video player or as a video streaming receiver.

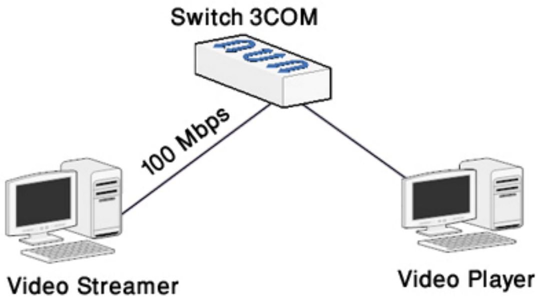


Figure 4. Implementation architecture

The main idea from video streaming implementation is streaming video with different protocols and different codecs. Then the performance of Mansoura University Network will be measured by using the following parameters:

- **Delay time**

$$\textbf{delay} = d_{proc} + d_{queue} + d_{trans} + d_{prop} \qquad \textbf{(1)}$$

d_{proc} process delay, d_{queue} queue delay, d_{trans} transmission delay, and d_{prop} propagation delay.

- **Jitter time**

$$\textbf{jitter (i) = delay(i+1) – delay(i)} \qquad \textbf{(2)}$$

When Server receives N packets from client, i = 1 to N.

- **Overall bandwidth**

$$\text{Overall bandwidth} = \sum_{i=1}^{N} \frac{(P\ length - P\ headers) * 8}{di} \qquad \textbf{(3)}$$

- **Average bandwidth**

$$\text{Avarage bandwidth} = \frac{1}{N}\sum_{i=1}^{N} \frac{(P\ length - P\ headers) * 8}{J(i)} \qquad \textbf{(4)}$$

- **Instance bandwidth**

$$\text{Instance bandwidth} = \frac{(P\ length - P\ headers) * 8}{J(i)} \qquad \textbf{(5)}$$

P $_{\textbf{length}}$ packet length, **P** $_{\textbf{header}}$ packet headers. **J(i)** jitter, **d**$_{\textbf{i}}$ dealy time

From the previous equations 1,2,3,4, and 5 the following figures are the results. The results will be used to select the protocol and codec combination for getting the best performance of streaming.

In this experimental used QoS of streaming video traffic guidelines as the percent of packets loss should be no more than 5%, latency should be no more than 4 to 5 seconds, bandwidth requirements depend on the encoding format and rate of the video stream should be guaranteed [10].

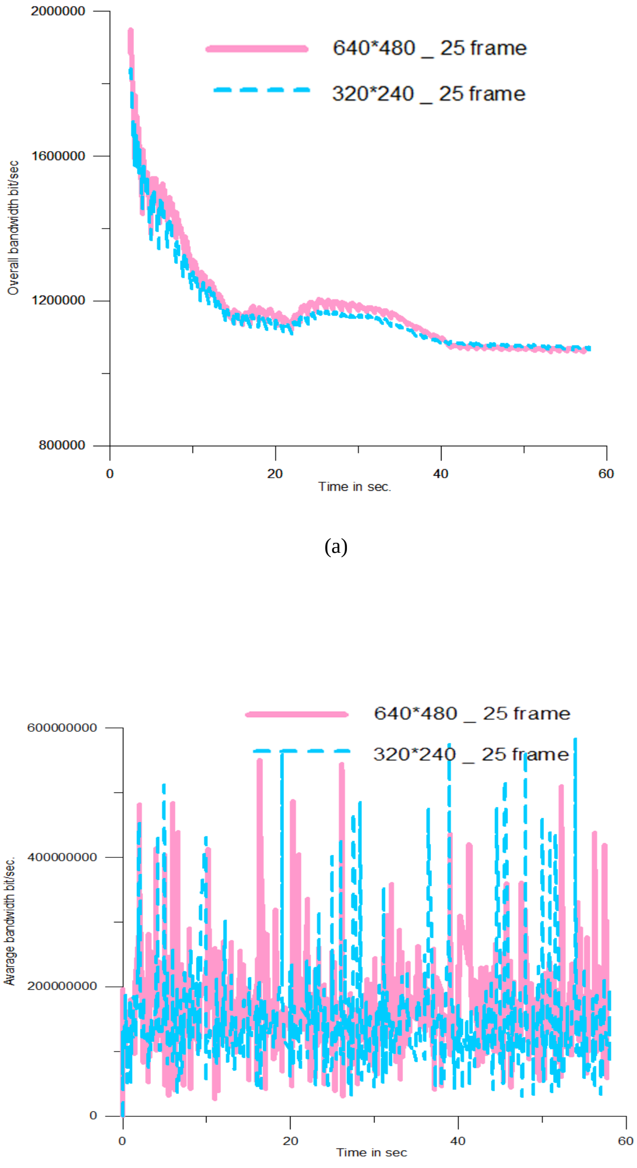
The results of first case, first scenario HTTP protocol with H.264 codec, with videos frame size 320*240 and 640*480, 25 frames is shown in Fig. 5.

The results of first case, second scenario HTTP protocol with MPEG-2 codec, with videos frame size 320*240 and 640*480, 25 frames is shown in Fig.6.

The results of second case, first scenario UDP protocol with H.264 codec, with videos frame size 320*240 and 640*480, 25 frames is shown in Fig.7.

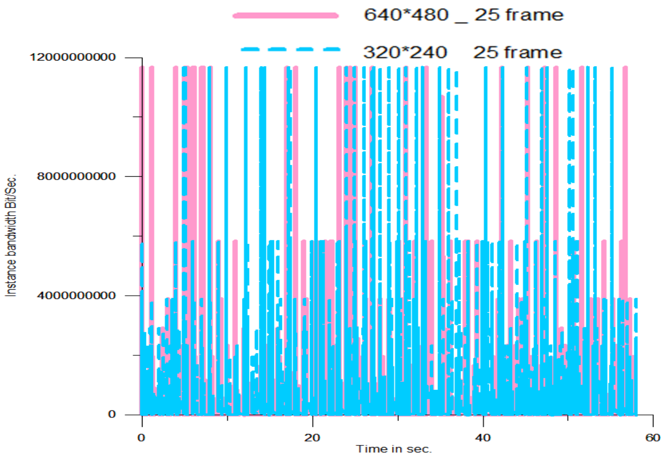
The results of second case, first scenario UDP protocol with MPEG-2 codec, with videos frame size 320*240 and 640*480, 25 frames is shown in Fig.8.

The compassion between video before sending and video after receiving is in Tables III, IV respectively.

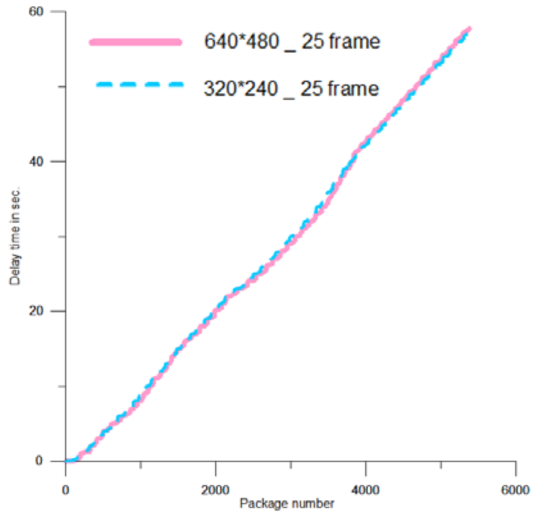


(a)

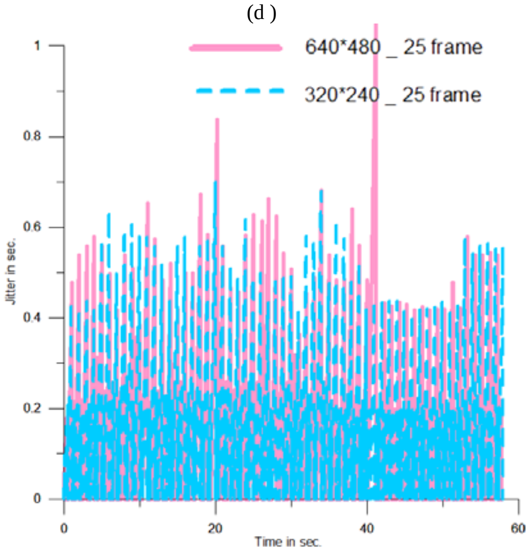
(b)



(c)

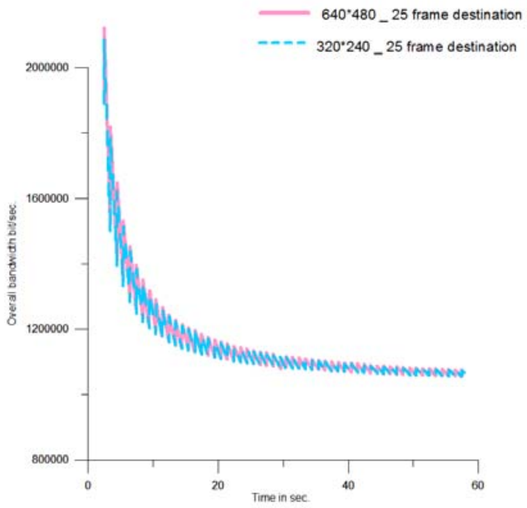


(d)

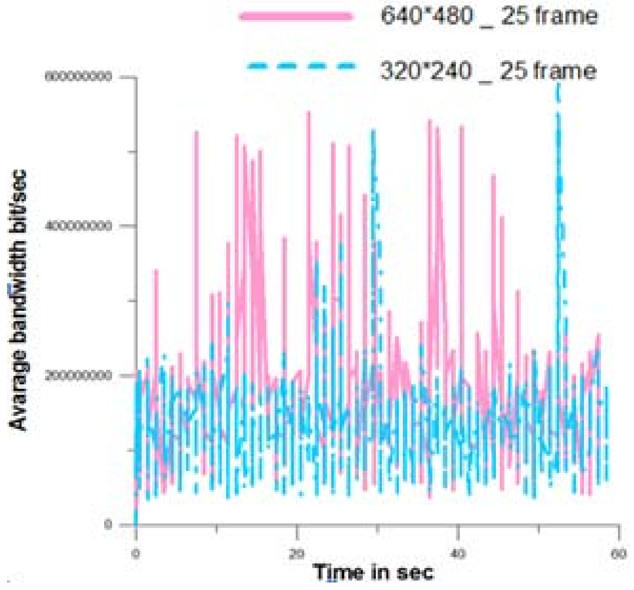


(e)

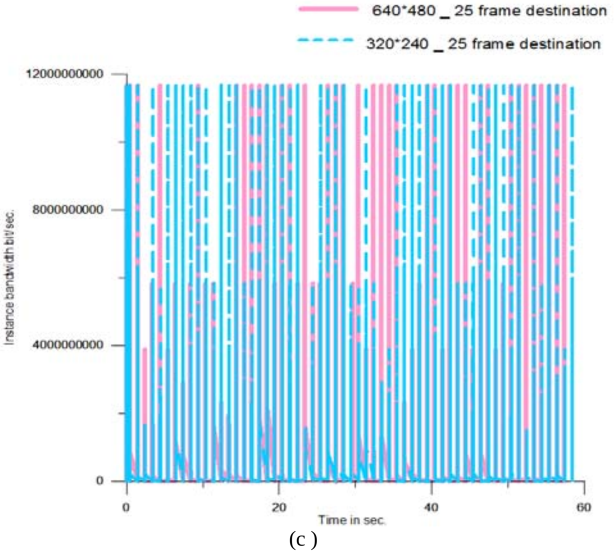
Figure 5. Network measurement parameters for H.264 over HTTP protocol. (a) overall bandwidth (b) average bandwidth (c) instance bandwidth (d) delay time (e) jitter time



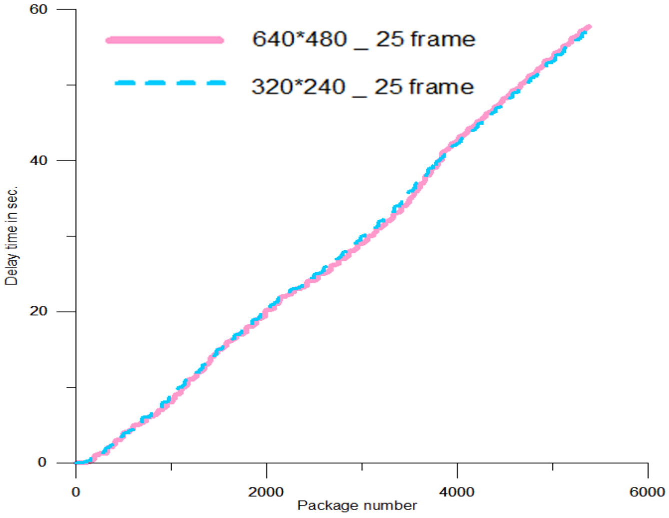
(a)



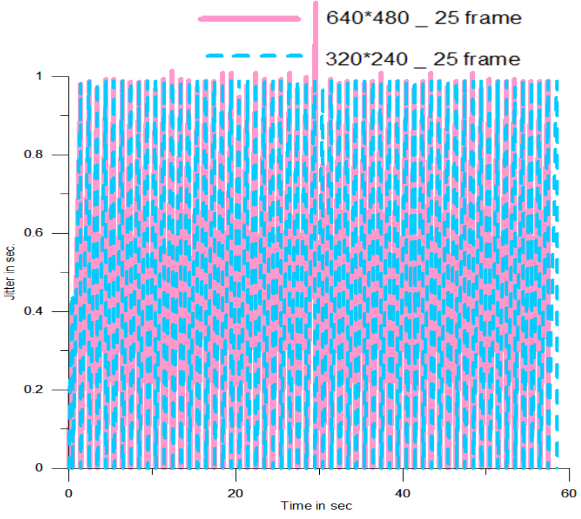
(b)



(c)

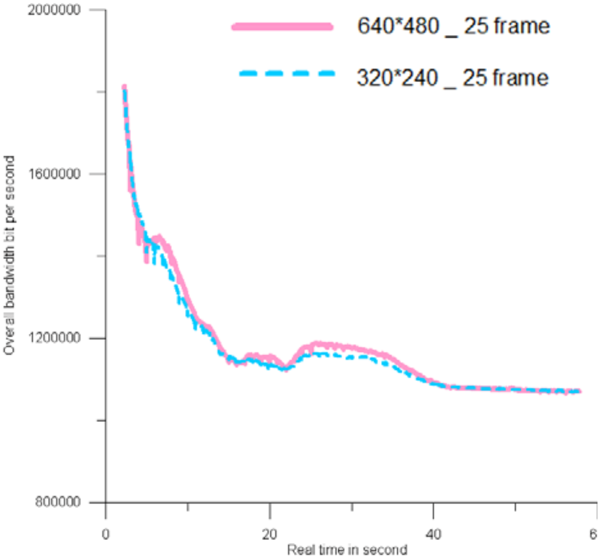


(d)

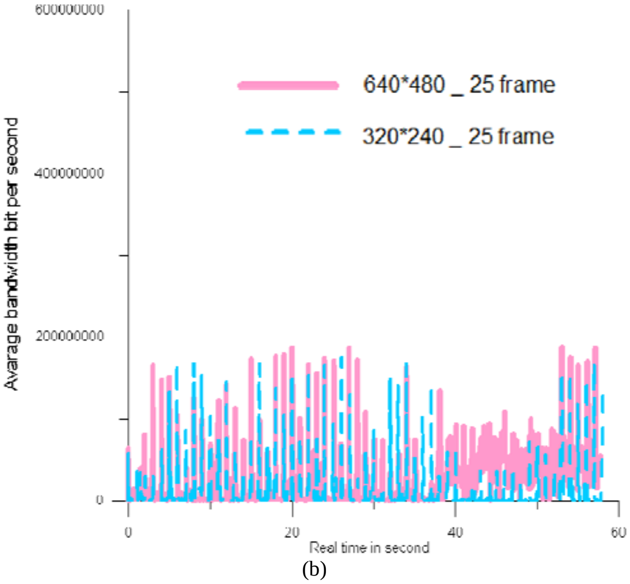


(e)

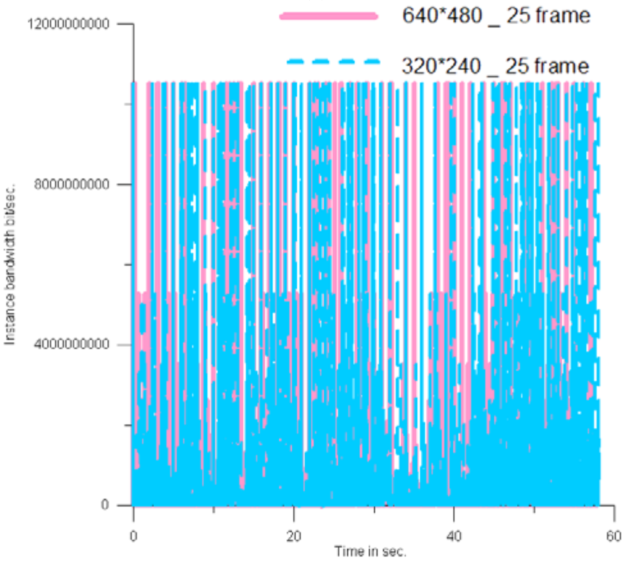
Figure 6. Network measurement parameters for MPEG-2 over HTTP protocol (a) overall bandwidth (b) average bandwidth (c) instance bandwidth (d) delay time (e) jitter time



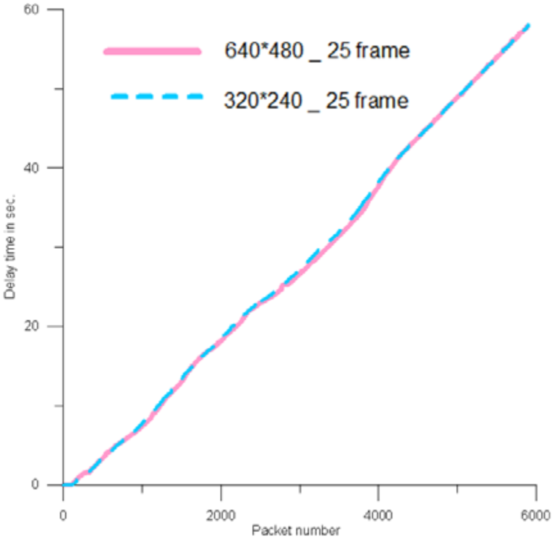
(a)



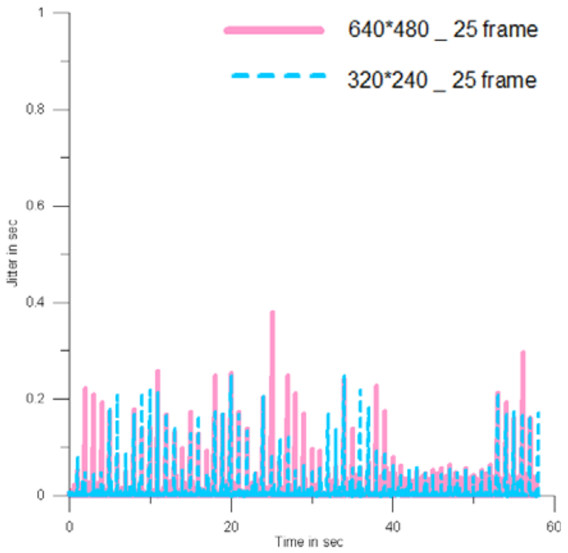
(b)



(c)

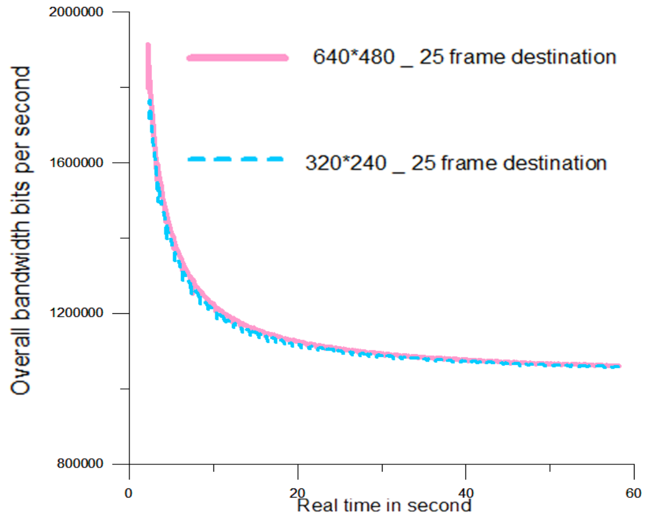


(d)

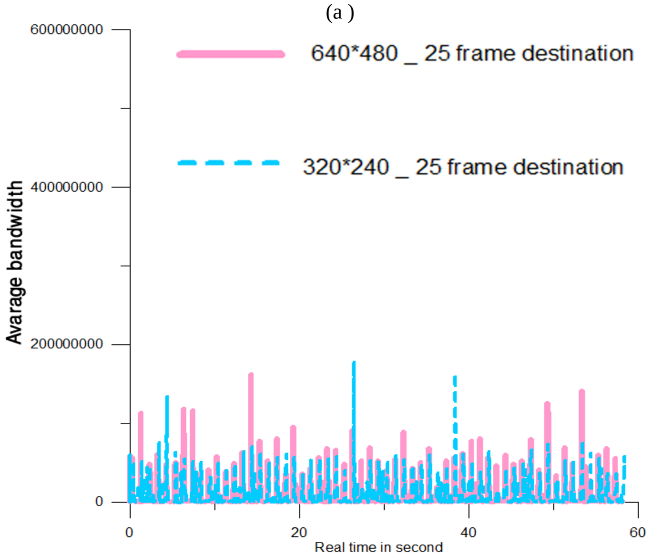


(e)

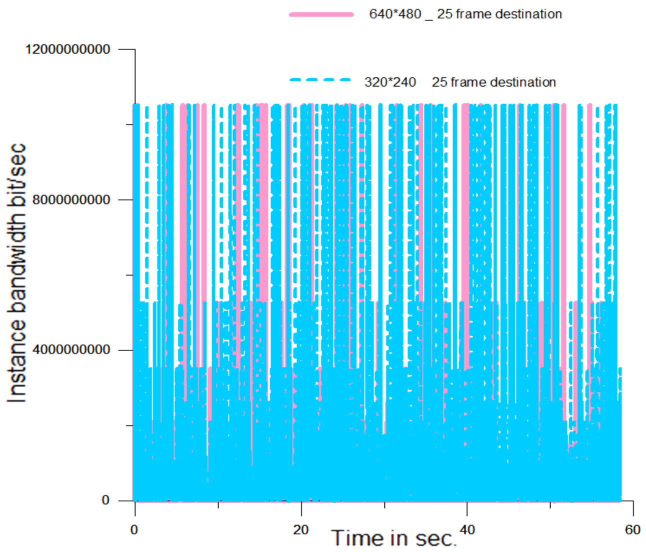
Figure 7. Network measurement parameters for H.264 over UDP protocol (a) overall bandwidth (b) average bandwidth (c) instance bandwidth (d) delay time (e) jitter time



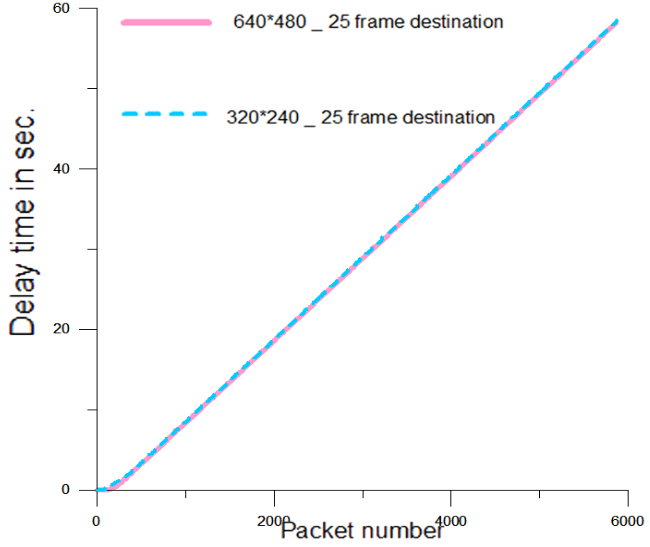
(a)



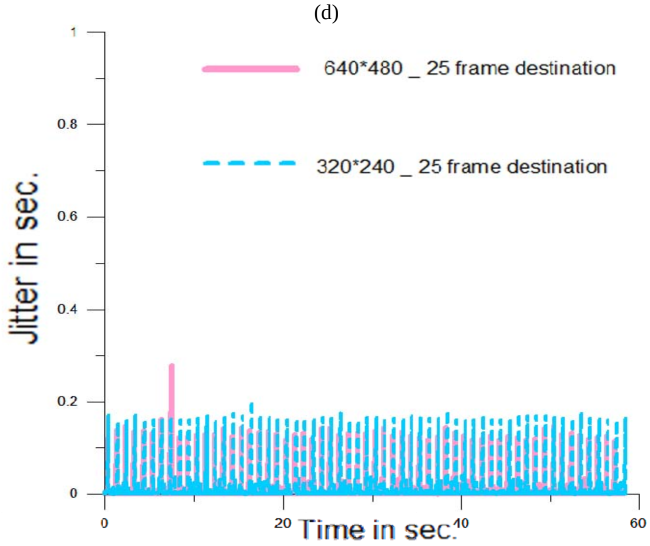
(b)



(c)



(d)



(e)

Figure 8. Network measurement parameters for MPEG-2 over UDP protocol (a) overall bandwidth (b) average bandwidth (c) instance bandwidth (d) delay time (e) jitter time

V. Simulation Results

The results show that The MPEG-2 is the best codec over UDP protocol, and H.264 is the best one over HTTP protocol. But in the view of video quality and calculate the loss of packets in each case.

MPEG-2 is better than H.264, because the percentage of packets loss is less than H.264 percentage. As described in tables V,VI.

QoS needs of Streaming-Video traffic, the following guidelines are recommended:

1. Loss should be no more than 5 percent.
2. Latency should be no more than 4 to 5 seconds (depending on the video application's buffering capabilities).
3. There are no significant jitter requirements.
4. Guaranteed bandwidth requirements depend on the encoding format and rate of the video stream.

From the previous guidelines and Tables IV,V. The MPEG-2 achieved these guides, but H.264 not recommended all guides recommended

IV. Conclusion

In this paper, the transmission of video streaming over the computer networks has been studied. This study was been performed on two different codecs H.264 and MPEG-2. Video streaming files have been transmitted by using two different protocols HTTP and UDP. The performance of transmission was been measured over Mansoura University computers network in real time. Practical results have shown that jitter time of MPEG-2 is less than H.264. So, MPEG-2 protocol is better than H.264 over the UDP protocols. In contrast, jitter time of H.264 is less than MPEG-2 over HTTP protocol. So, H.264 is better than MPEG-2 over the HTTP protocol. This is from the network performance view. However, from video quality view, MPEG-2 has achieved the guidelines of QoS of video streaming.

References

[1] Jerry D. Gibson, *Multimedia Communications Directions and Innovations*, Academic Pres, 2001

[2]Noriaki Kamiyama, Ryoichi Kawahara, Tatsuya Mori, Shigeaki Harada, Haruhisa Hasegawa, "Parallel video streaming optimizing network throughput", Elsevier, 2010.

[3]Tarek R. Sheltami, Elhadi M. Shakshuki, Hussein T. Mouftah, " Video streaming application over WEAC protocol in MANET ", Elsevier, 2010.

[4]Yung-Sung Huang, "A portable medical system using real-time streaming transport over 3G wireless networks", Springer-Verlag, 2010

[5] Iain E.G. Richardson, *Video Codec Design- Developing Image and Video Compression Systems*, John Wiley & Sons Ltd, 2002.

[6] Jörn Ostermann, Jan Bormans, Peter List, Detlev Marpe, Matthias Narroschke, Fernando Pereira, Thomas Stockhammer, and Thomas Wedi, *Video coding with H.264/AVC: Tools, Performance, and Complexity*, IEEE CIRCUITS AND SYSTEMS MAGAZINE, 2004.

[7] K. R. Rao, Zoran S. Bojkovic, Dragorad A. Milovanovic, *Introduction to multimedia communications Applications, Middleware, Networking*, JOHN WILEY & SONS, INC., 2006.

[8] Gorry Fairhust, "*MPEG-2*", http://web.archive.org/web/20071023065124/http://erg.abdn.ac.uk/research/future-net/digital-video/mpeg2.html , 2001.

[9] ISO/IEC FCD 14496, "*Information technology – Coding of audio-visual objects – Part 2: Visual,*" July 2001.

[10] Tim Szigeti-Christina Hattingh, *End-to-End QoS Network Design*, Cisco Press, November 09, 2004

Table III. Comparison between video frame size 320*240 with 25 frames before sending and after receiving

Video size 320_240_25f send	Video size 320_240_25f received compressed HTTP and H.264	Video size 320_240_25f received compressed UDP and H.264	Video size 320_240_25f received compressed HTTP and MPEG-2	Video size 320_240_25f received compressed UDP and MPEG-2
Filename: '320_240_25f_send.avi' FileSize: 357399868 NumFrames: 1500 FramesPerSecond: 25 Width: 320 Height: 240 ImageType: 'truecolor' VideoCompression: 'none' Quality: 0 NumColormapEntries: 0 AudioFormat: 'PCM' AudioRate: 48000 NumAudioChannels: 2	Filename: '320_240_25f_http_h264.avi' FileSize: 5704948 NumFrames: 1401 FramesPerSecond: 23.9070 Width: 320 Height: 240 ImageType: 'truecolor' VideoCompression: 'h264' Quality: 4.2950e+007 NumColormapEntries: 0	Filename: '320_240_25f_udp_h264.avi' FileSize: 5656754 NumFrames: 1415 FramesPerSecond: 23.9280 Width: 320 Height: 240 ImageType: 'truecolor' VideoCompression: 'h264' Quality: 4.2950e+007 NumColormapEntries: 0	Filename: '320_240_25f_http_mpeg2.avi' FileSize: 7186994 NumFrames: 1493 FramesPerSecond: 25 Width: 320 Height: 240 ImageType: 'truecolor' VideoCompression: 'mpgv' Quality: 4.2950e+007 NumColormapEntries: 0 AudioFormat: 'Format # 0x55' AudioRate: 44100 NumAudioChannels: 2	Filename: '320_240_25f_udp_mpeg2.avi' FileSize: 7091670 NumFrames: 1476 FramesPerSecond: 25.0220 Width: 320 Height: 240 ImageType: 'truecolor' VideoCompression: 'mpgv' Quality: 4.2950e+007 NumColormapEntries: 0 AudioFormat: 'Format # 0x55' AudioRate: 44100 NumAudioChannels: 2

Table IV. Comparison between video frame size 640*480 with 25 frames before sending and after receiving

Video size 640_480_25f send uncompressed	Video size 640_480_25f received compressed HTTP and H.264	Video size 640_480_25f received compressed UDP and H.264	Video size 640_480_25f received compressed HTTP and MPEG-2	Video size 640_480_25f received compressed UDP and MPEG-2
Filename: '640_480_25f_send.avi' FileSize: 37329305 NumFrames: 1500 FramesPerSecond: 25 Width: 640 Height: 480 ImageType: 'truecolor' VideoCompression: 'none' Quality: 0 NumColormapEntries: 0 AudioFormat: 'PCM' AudioRate: 48000 NumAudioChannels: 2	Filename: '640_480_25f_http_h264.avi' FileSize: 5672488 NumFrames: 1392 FramesPerSecond: 21.2110 Width: 640 Height: 480 ImageType: 'truecolor' VideoCompression: 'h264' Quality: 4.2950e+007 NumColormapEntries: 0	Filename: '640_480_25f_udp_h264_.avi' FileSize: 5452408 NumFrames: 1372 FramesPerSecond: 20.5570 Width: 640 Height: 480 ImageType: 'truecolor' VideoCompression: 'h264' Quality: 4.2950e+007 NumColormapEntries: 0	Filename: '640_480_25f_http_mpeg2.avi' FileSize: 7069402 NumFrames: 1469 FramesPerSecond: 27.7280 Width: 640 Height: 480 ImageType: 'truecolor' VideoCompression: 'mpgv' Quality: 4.2950e+007 NumColormapEntries: 0 AudioFormat: 'Format # 0x55' AudioRate: 44100 NumAudioChannels: 2	Filename: '640_480_25f_udp_mpeg2.avi' FileSize: 7067402 NumFrames: 1452 FramesPerSecond: 27.7280 Width: 640 Height: 480 ImageType: 'truecolor' VideoCompression: 'mpgv' Quality: 4.2950e+007 NumColormapEntries: 0 AudioFormat: 'Format # 0x55' AudioRate: 44100 NumAudioChannels: 2

Table V. Jitter time comparison between H.264 and MPEG-2 over UDP and HTTP

	MPEG-2	H .264	
UDP	Jitter is less than 0.2 second	Jitter is less than 0.25 second	320_240_25f
	Jitter is less than 0.14 second	Jitter is less than 0.25 second	320_240_30f
	Jitter is really less than 0.2 sec 0.3	Jitter is really less than 0.3 sec 0.4	640_480_25f
	Jitter is really less than 0.2 sec 1.4	Jitter is really less than 0.2 sec 4	640_480_30f
HTTP	Jitter is less than 1 sec	Jitter is less than 0.8 sec	320_480_25f
	Jitter is less than 1 sec	Jitter is less than 0.8 sec	320_480_30f
	Jitter is really less than 1 sec 1.2	Jitter is really less than 1 sec 1.2	640_480_25f
	Jitter is really less than 1 sec 1.6	Jitter is really less than 1 sec 1.3	640_480_30f

Table VI. Packets loss percentage comparison between H.264 and MPEG-2 over UDP and HTTP

	Mpeg	H .264	
UDP	Packets loss is : 1.6%	Packets loss is : 5.7%	320_240_25f
	Packets loss is :2.5%	Packets loss is :6.1%	320_240_30f
	Packets loss is :3.2%	Packets loss is :8.5%	640_480_25f
	Packets loss is :4.7%	Packets loss is :6.1%	640_480_30f
HTTP	Packets loss is :0.5%	Packets loss is : 6.6%	320_480_25f
	Packets loss is :0.4%	Packets loss is :5.7%	320_480_30f
	Packets loss is :2.1%	Packets loss is :7.2%	640_480_25f
	Packets loss is :1.2%	Packets loss is :5.9%	640_480_30f

Fast Detection of H1N1 and H1N5 Viruses in DNA Sequence by using High Speed Time Delay Neural Networks

Hazem M. El-Bakry
Faculty of Computer Science & Information Systems,
Mansoura University, EGYPT
helbakry20@yahoo.com

Nikos Mastorakis
Technical University of Sofia,
BULGARIA

Abstract—**Fast detection of biological viruses in DNA sequence is very important for investigation of patients and overcome diseases. First, an intelligent algorithm to completely retrieve DNA sequence is presented. DNA codes that may be missed during the splitting process are retrieved by using Hopfield neural networks. Then, a new approach for fast detection of biological viruses like H1N1 and H1N5 in DNA sequence is presented. Such algorithm uses high speed time delay neural networks (HSTDNNs). The operation of these networks relies on performing cross correlation in the frequency domain between the input DNA sequence and the input weights of neural networks. It is proved mathematically and practically that the number of computation steps required for the presented HSTDNNs is less than that needed by conventional time delay neural networks (CTDNNs). Simulation results using MATLAB confirm the theoretical computations.**

Keywords- High Speed Neural Networks; Cross Correlation; Frequency Domain; H1N1 and H1N5 Detection

I. INTRODUCTION

A virus is a tiny bundle of genetic material - either DNA or RNA - carried in a shell called a viral coat, or capsid, which is made up of protein. Some viruses have an additional layer around this coat called an envelope. When a virus particle enters a cell and begins to reproduce itself, this is called a viral infection. The virus is usually very, very small compared to the size of a living cell. The information carried in the virus's DNA allows it to take over the operation of the cell, converting it to a factory to make more copies of itself. For example, the polio virus can make over one million copies of itself inside a single, infected human intestinal cell [32-35].

All viruses only exist to make more viruses. With the possible exception of bacterial viruses, which can kill harmful bacteria, all viruses are considered harmful, because their reproduction causes the death of the cells which the viruses entered. If a virus contains DNA, it inserts its genetic material into the host cell's DNA. If the virus contains RNA, it must first turn its RNA into DNA using the host cell's machinery, before inserting it into the host DNA. Once it has taken over the cell, viral genes are then copied thousands of times, using the machinery the host cell would ordinarily use to reproduce its own DNA. Then the host cell is forced to encapsulate this viral DNA into new protein shells; the new viruses created are then released, destroying the cell [32-35].

All living things are susceptible to viral infections plants, animals, or bacteria can all be infected by a virus specific for that type of organism. Moreover, within an individual species there may be a hundred or more different viruses which can infect that species alone. There are viruses which infect only humans (for example, smallpox), viruses which infect humans and one or two additional kinds of animals (for example, influenza), viruses which infect only a certain kind of plant (for example, the tobacco mosaic virus), and some viruses which infect only a particular species of bacteria (for example, the bacteriophage which infects E. coli) [32-35].

Sometimes when a virus reproduces, mutations occur. The offspring that have been changed by the mutation may no longer be infectious. But a virus replicates itself thousands of times, so there will usually be some offspring that are still infectious, but sufficiently different from the parent virus so that vaccines no longer work to kill it. The influeza virus can do this, which is why flu vaccines for last year's flu don't work the next year. The common cold virus changes so quickly that vaccines are useless; the cold you have today will be a different strain than the cold you had last month! [31-34]

For efficient treatment of patients in real-time, it is important to detect biological viruses like H1N1 and H1N5. Recently, time delay neural networks have shown very good results in different areas such as automatic control, speech recognition, blind equalization of time-varying channel and other communication applications. The main objective of this research is to reduce the response time of time delay neural networks. The purpose is to perform the testing process in the frequency domain instead of the time domain. Our approach was successfully applied for fast detection of computer viruses as shown in [4]. Sub-image detection by using fast neural networks (FNNs) was proposed in [5,6]. Furthermore, it was used for fast face detection [7,10,12], and fast iris detection

[11]. Another idea to further increase the speed of FNNs through image decomposition was suggested in [10]. In addition it was applied for fast prediction of new data as described in [1,3].

FNNs for detecting a certain code in one dimensional serial stream of sequential data were described in [1,2,3,4,8,14,15,20,23,27,28,29]. Compared with conventional neural networks, FNNs based on cross correlation between the tested data and the input weights of neural networks in the frequency domain showed a significant reduction in the number of computation steps required for certain data detection [1-29]. Here, we make use of the theory of FNNs implemented in the frequency domain to increase the speed of time delay neural networks for biological virus detection [2]. The idea of moving the testing process from the time domain to the frequency domain is applied to time delay neural networks. Theoretical and practical results show that the proposed HSTDNNs are faster than CTDNNs. Retrieval of missed DNA codes by using Hopfield neural networks is introduced in section II. Section III presents HSTDNNs for detecting of biological viruses in DNA sequence. Experimental results for fast biological virus detection by using HSTDNNs are given in section IV.

II. Retrieval of Missed DNA Codes by using Hopfield Neural Networks

One of the most important functions of our brain is the laying down and recall of memories. It is difficult to imagine how we could function without both short and long term memory. The absence of short term memory would render most tasks extremely difficult if not impossible - life would be punctuated by a series of one time images with no logical connection between them. Equally, the absence of any means of long term memory would ensure that we could not learn by past experience. Indeed, much of our impression of self depends on remembering our past history [36-40].

Our memories function in what is called an associative or content-addressable fashion. That is, a memory does not exist in some isolated fashion, located in a particular set of neurons. All memories are in some sense strings of memories - you remember someone in a variety of ways - by the color of their hair or eyes, the shape of their nose, their height, the sound of their voice, or perhaps by the smell of a favorite perfume. Thus memories are stored in *association* with one another. These different sensory units lie in completely separate parts of the brain, so it is clear that the memory of the person must be distributed throughout the brain in some fashion. Indeed, PET scans reveal that during memory recall there is a pattern of brain activity in many widely different parts of the brain [36-43].

Notice also that it is possible to access the full memory (all aspects of the person's description for example) by initially remembering just one or two of these characteristic features. We access the memory by its contents not by where it is stored in the neural pathways of the brain. This is very powerful; given even a poor photograph of that person we are quite good at reconstructing the persons face quite accurately. This is very different from a traditional computer where specific facts are located in specific places in computer memory. If only partial information is available about this location, the fact or memory cannot be recalled at all [35-42].

Theoretical physicists are an unusual lot, acting like gunslingers in the old West, anxious to prove themselves against a really good problem. And there aren't that many really good problems that might be solvable. As soon as Hopfield pointed out the connection between a new and important problem (network models of brain function) and an old and well-studied problem (the Ising model), many physicists rode into town, so to speak, with the intention of shooting the problem full of holes and then, the brain understood, riding off into the sunset looking for a newer, tougher problem. (Who was that masked physicist?).

Hopfield made the portentous comment: 'This case is isomorphic with an Ising model,' thereby allowing a deluge of physical theory (and physicists) to enter neural network modeling. This flood of new participants transformed the field. In 1974 Little and Shaw made a similar identification of neural network dynamics with the Ising model, but for whatever reason, their idea was not widely picked up at the time. Unfortunately, the problem of brain function turned out to be more difficult than expected, and it is still unsolved, although a number of interesting results about Hopfield nets were proved. At present, many of the traveling theoreticians have traveled on [38].

The Hopfield neural network is a simple artificial network which is able to store certain memories or patterns in a manner rather similar to the brain - the full pattern can be recovered if the network is presented with only partial information. Furthermore there is a degree of stability in the system - if just a few of the connections between nodes (neurons) are severed, the recalled memory is not too badly corrupted - the network can respond with a "best guess". Of course, a similar phenomenon is observed with the brain - during an average lifetime many neurons will die but we do not suffer a catastrophic loss of individual memories - our brains are quite robust in this respect (by the time we die we may have lost 20 percent of our original neurons) [44-57].

The nodes in the network are vast simplifications of real neurons - they can only exist in one of two possible "states" - firing or not firing. Every node is connected to every other node with some strength. At any instant of time a node will change its state (i.e start or stop firing) depending on the inputs it receives from the other nodes [44-57].

If we start the system off with a any general pattern of firing and non-firing nodes then this pattern will in general change with time. To see this think of starting the network with just one firing node. This will send a signal to all the other nodes via its connections so that a short time later some of these other nodes will fire. These new firing nodes will then excite others after a further short time interval and a whole cascade of different firing patterns will occur. One might imagine that the firing pattern of the network would change in a complicated perhaps random way with time. The crucial property of the Hopfield network which renders it useful for simulating memory recall is the following: we are guaranteed that the pattern will settle down after a long enough time to some fixed pattern. Certain nodes will be always "on" and others "off". Furthermore, it is possible to arrange that these stable firing patterns of the network correspond to the desired memories we wish to store! [44-57].

The reason for this is somewhat technical but we can proceed by analogy. Imagine a ball rolling on some bumpy surface. We imagine the position of the ball at any instant to represent the activity of the nodes in the network. Memories will be represented by special patterns of node activity corresponding to wells in the surface. Thus, if the ball is let go, it will execute some complicated motion but we are certain that eventually it will end up in one of the wells of the surface. We can think of the height of the surface as representing the energy of the ball. We know that the ball will seek to minimize its energy by seeking out the lowest spots on the surface -- the wells. Furthermore, the well it ends up in will usually be the one it started off closest to. In the language of memory recall, if we start the network off with a pattern of firing which approximates one of the "stable firing patterns" (memories) it will "under its own steam" end up in the nearby well in the energy surface thereby recalling the original perfect memory. The smart thing about the Hopfield network is that there exists a rather simple way of setting up the connections between nodes in such a way that any desired set of patterns can be made "stable firing patterns". Thus any set of memories can be burned into the network at the beginning. Then if we kick the network off with any old set of node activity we are *guaranteed* that a "memory" will be recalled. Not too surprisingly, the memory that is recalled is the one which is "closest" to the starting pattern. In other words, we can give the network a corrupted image or memory and the network will "all by itself" try to reconstruct the perfect image. Of course, if the input image is sufficiently poor, it may recall the incorrect memory - the network can become "confused" - just like the human brain. We know that when we try to remember someone's telephone number we will sometimes produce the wrong one! Notice also that the network is reasonably robust - if we change a few connection strengths just a little the recalled images are "roughly right". We don't lose any of the images completely [44-57].

As with the Linear Associative Memory, the “stored patterns” are represented by the weights. To be effective, the patterns should be reasonably orthogonal. The basic Hopfield model can be described as follows [38]:

- N neurons, fully connected in a cyclic fashion:
- Values are +1, -1.
- Each neuron has a weighted input from all other neurons.
- The weight matrix w is symmetric ($w_{ij=}w_{ji}$) and diagonal terms (self-weights $w_{ii} = 0$).
- Activation function on each neuron i is:

$$f(\text{net}) = \text{sgn}(\text{net}) = \begin{Bmatrix} 1 \text{ if net} > 0 \\ -1 \text{ if net} < 0 \end{Bmatrix} \quad (1)$$

where:

$$\text{net}_i = \Sigma w_{ij}\, x_j \quad (2)$$

- If net = 0, then the output is the same as before, by convention.
- There are no separate thresholds or biases. However, these could be represented by units that have all weights = 0 and thus never change their output.
- The energy function is defined as:

$$E(y_1, y_2, \ldots, y_n) = - \Sigma\, \Sigma\, w_{ij}\, y_i y_j \quad (3)$$

where $(y_1, y_2, \ldots, y_n)$ is outputs, w_{ij} is the weight neuron i, and the double sum is over i and j.

Different DNA patterns are stored in Hopfield neural network. In the testing process, the missed codes (if any) are retrieved.

III. Fast Biological Virus Detection by using HSTDNNs

Finding a biological virus like H1N1 or H1N5 in DNA sequence is a searching problem. First neural networks are trained to classify codes which contain viruses from others that do not and this is done in time domain. In biological virus detection phase, each position in the DNA sequence is tested for presence or absence of biological virus code. At each position in the input DNA one dimensional matrix, each sub-matrix is multiplied by a window of weights, which has the same size as the sub-matrix. The outputs of neurons in the hidden layer are multiplied by the weights of the output layer. When the final output is 10, this means that the sub-matrix under test contains H1N1. When the final output is 01 this means that H1N5 is detected. Otherwise, there is no virus. Thus, we may conclude that this searching problem is a cross correlation between the incoming serial data and the weights of neurons in the hidden layer.

The convolution theorem in mathematical analysis says that a convolution of f with h is identical to the result of the following steps: let F and H be the results of the Fourier Transformation of f and h in the frequency domain. Multiply F and H* in the frequency domain point by point and then transform this product into the spatial domain via the inverse Fourier Transform. As a result, these cross correlations can be represented by a product in the frequency domain. Thus, by using cross correlation in the frequency domain, speed up in an order of magnitude can be achieved during the detection process [1-29]. Assume that the size of the biological virus code is 1xn. In biological virus detection phase, a sub matrix I of size 1xn (sliding window) is extracted from the tested matrix, which has a size of 1xN. Such sub matrix, which may be biological virus code, is fed to the neural network. Let W_i be the matrix of weights between the input sub-matrix and the hidden layer. This vector has a size of 1xn and can be represented as 1xn matrix. The output of hidden neurons h(i) can be calculated as follows [1-7]:

$$h_i = g\left(\sum_{k=1}^{n} W_i(k) I(k) + b_i\right) \qquad (4)$$

where g is the activation function and b(i) is the bias of each hidden neuron (i). Equation 4 represents the output of each hidden neuron for a particular sub-matrix I. It can be obtained to the whole input matrix Z as follows [1-6]:

$$h_i(u)=g\left(\sum_{k=-n/2}^{n/2} W_i(k)\ Z(u+k)+b_i\right) \qquad (5)$$

Eq.5 represents a cross correlation operation. Given any two functions f and d, their cross correlation can be obtained by [31]:

$$d(x)\otimes f(x)=\left(\sum_{n=-\infty}^{\infty} f(x+n)d(n)\right) \qquad (6)$$

Therefore, Eq. 5 may be written as follows [1-7]:

$$h_i = g\left(W_i \otimes Z + b_i\right) \qquad (7)$$

where h_i is the output of the hidden neuron (i) and h_i (u) is the activity of the hidden unit (i) when the sliding window is located at position (u) and (u) $\in$ [N-n+1].

Now, the above cross correlation can be expressed in terms of one dimensional Fast Fourier Transform as follows [1-7]:

$$W_i \otimes Z = F^{-1}\left(F(Z)\bullet F^*\left(W_i\right)\right) \qquad (8)$$

Hence, by evaluating this cross correlation, a speed up ratio can be obtained comparable to conventional neural networks. Also, the final output of the neural network can be evaluated as follows:

$$O(u) = g\left(\sum_{i=1}^{q} W_o(i)\, h_i(u) + b_o\right) \qquad (9)$$

where q is the number of neurons in the hidden layer. O(u) is the output 2D matrix (corresponding to two output neurons) of the neural network when the sliding window located at the position (u) in the input matrix Z. W_o is the weight matrix between hidden and output layer.

IV. Complexity Analysis of HSTDNNs for Biological Virus Detection

The complexity of cross correlation in the frequency domain can be analyzed as follows:

1- For a tested matrix of 1xN elements, the 1D-FFT requires a number equal to $N\log_2 N$ of complex computation steps [30]. Also, the same number of complex computation steps is required for computing the 1D-FFT of the weight matrix at each neuron in the hidden layer.

2- At each neuron in the hidden layer, the inverse 1D-FFT is computed. Therefore, q backward and (1+q) forward transforms have to be computed. Therefore, for a given matrix under test, the total number of operations required to compute the 1D-FFT is $(2q+1)N\log_2 N$.

3- The number of computation steps required by HSTDNNs is complex and must be converted into a real version. It is known that, the one dimensional Fast Fourier Transform requires $(N/2)\log_2 N$ complex multiplications and $N\log_2 N$ complex additions [30]. Every complex multiplication is realized by six real floating point operations and every complex addition is implemented by two real floating point operations. Therefore, the total number of computation steps required to obtain the 1D-FFT of a 1xN matrix is:

$$\rho=6((N/2)\log_2 N) + 2(N\log_2 N) \qquad (10)$$

which may be simplified to:

$$\rho=5N\log_2 N \qquad (11)$$

4- Both the input and the weight matrices should be dot multiplied in the frequency domain. Thus, a number of complex computation steps equal to qN should be considered. This means 6qN real operations will be added to the number of computation steps required by HSTDNNs.

5- In order to perform cross correlation in the frequency domain, the weight matrix must be extended to have the same size as the input matrix. So, a number of zeros = (N-n) must be added to the weight matrix. This requires a total real number of computation steps = q(N-n) for all neurons. Moreover, after computing the FFT for the weight matrix, the conjugate of this matrix must be obtained. As a result, a real number of computation steps = qN should be added in order to obtain the conjugate of the weight matrix for all neurons. Also, a number of real computation steps equal to N is required to create butterflies complex numbers ($e^{-jk(2\Pi n/N)}$), where 0<K<L. These (N/2) complex numbers are multiplied by the elements of the input matrix or by previous complex numbers during the computation of FFT. To create a complex number requires two real floating point operations. Thus, the total number of computation steps required for HSTDNNs becomes:

$$\sigma=(2q+1)(5N\log_2 N)+6qN+q(N-n)+qN+N \qquad (12)$$

which can be reformulated as:

$$\sigma=(2q+1)(5N\log_2 N)+q(8N-n)+N \qquad (13)$$

6- Using sliding window of size 1xn for the same matrix of 1xN pixels, q(2n-1)(N-n+1) computation steps are required when using CTDNNs for biological virus detection or processing (n) input data. The theoretical speed up factor η can be evaluated as follows:

$$\eta = \frac{q(2n-1)(N-n+1)}{(2q+1)(5N\log_2 N)+q(8N-n)+N} \qquad (14)$$

CTDNNs and HSTDNNs are shown in Figures 1 and 2 respectively.

Time delay neural networks accept serial input data with fixed size (n). Therefore, the number of input neurons equals to (n). Instead of treating (n) inputs, the proposed new approach is to collect all the incoming data together in a long vector (for example 100xn). Then the input data is tested by time delay neural networks as a single pattern with length L (L=100xn). Such a test is performed in the frequency domain as described before.

The theoretical speed up ratio for searching short successive (n) code in a long input vector (L) using time delay neural networks is listed in tables I, II, and III. Also, the practical speed up ratio for manipulating matrices of different sizes (L) and different sized weight matrices (n) using a 2.7 GHz processor and MATLAB is shown in table IV.

An interesting point is that the memory capacity is reduced when using HSTDNN. This is because the number of variables is reduced compared with CTDNN.

V. Conclusion

To facilitate investigation of patients and overcome diseases, fast detection of biological viruses in DNA sequence has been presented. Missed DNA codes have been retrieved by using Hopfield neural networks. After that a new approach for fast detection of biological viruses like H1N1 and H1N5 in DNA sequence has been introduced. Such strategy has been realized by using our design for HSTDNNs. Theoretical computations have shown that HSTDNNs require fewer computation steps than conventional ones. This has been achieved by applying cross correlation in the frequency domain between the input data and the weights of neural networks. Simulation results have confirmed this proof by using MATLAB. The proposed algorithm can be applied to detect other biological viruses in DNA sequence perfectly.

References

[1] Hazem M. El-Bakry and Wael A. Awad, “A New Hybrid Neural Model for Real-Time Prediction Applications,” International Journal of Computer Science and Information Security, vol. 9, no. 5, May, 2011, pp. 244-255.

[2] Hazem M. El-Bakry, and Nikos Mastorakis, “An Intelligent Approach for Fast Detection of Biological Viruses in DNA Sequence,” Proc. of 10th WSEAS International Conference on APPLICATIONS of COMPUTER ENGINEERING (ACE '11), Spain, March 24-26, 2011, pp. 237-244.

[3] Hazem M. El-Bakry, and Nikos Mastorakis, “A New Approach for Prediction by using Integrated Neural Networks,” Proc. of 5th WSEAS International Conference on COMPUTER ENGINEERING and APPLICATIONS (CEA '11), Puerto Morelos, Mexico, Jan. 29-31, 2011, pp. 17-28.

[4] Hazem M. El-Bakry, "Fast Virus Detection by using High Speed Time Delay Neural Networks," Journal of Computer Virology, vol.6, no.2, 2010, pp.115-122.

[5] Hazem M. El-Bakry, "An Efficient Algorithm for Pattern Detection using Combined Classifiers and Data Fusion," Information Fusion Journal, vol. 11, 2010, pp. 133-148.

[6] Hazem M. El-Bakry, "A Novel High Speed Neural Model for Fast Pattern Recognition," Soft Computing Journal, vol. 14, no. 6, 2010, pp. 647-666.

[7] Hazem M. El-Bakry, "New Fast Principal Component Analysis For Real-Time Face Detection," MG&V Journal, vol. 18, no.4, 2009, pp. 405-426.

[8] Hazem M. El-bakry, and Mohamed Hamada “High speed time delay Neural Networks for Detecting DNA Coding Regions,” Springer, Lecture Notes on Artificial Intelligence (LNAI 5711), 2009, pp. 334-342.

[9] Hazem M. El-Bakry, "New Faster Normalized Neural Networks for Sub-Matrix Detection using Cross Correlation in the Frequency Domain and Matrix Decomposition, " Applied Soft Computing journal, vol. 8, issue 2, March 2008, pp. 1131-1149.

[10] Hazem M. El-Bakry, "Face detection using fast neural networks and image decomposition," Neurocomputing Journal, vol. 48, 2002, pp. 1039-1046.

[11] Hazem M. El-Bakry, "Human Iris Detection Using Fast Cooperative Modular Neural Nets and Image Decomposition," Machine Graphics & Vision Journal (MG&V), vol. 11, no. 4, 2002, pp. 498-512.

[12] Hazem M. El-Bakry, "Automatic Human Face Recognition Using Modular Neural Networks," Machine Graphics & Vision Journal (MG&V), vol. 10, no. 1, 2001, pp. 47-73.

[13] Hazem M. El-Bakry, "A New Neural Design for Faster Pattern Detection Using Cross Correlation and Matrix Decomposition," Neural World journal, Neural World Journal, 2009, vol. 19, no. 2, pp. 131-164.

[14] Hazem M. El-Bakry, and H. Stoyan, "FNNs for Code Detection in Sequential Data Using Neural Networks for Communication Applications," Proc. of the First International Conference on Cybernetics and Information Technologies, Systems and Applications: CITSA 2004, 21-25.

[15] Hazem M. El-Bakry, "New High speed time delay Neural Networks Using Cross Correlation Performed in the Frequency Domain," Neurocomputing Journal, vol. 69, October 2006, pp. 2360-2363.

[16] Hazem M. El-Bakry, "A New High Speed Neural Model For Character Recognition Using Cross Correlation and Matrix Decomposition," International Journal of Signal Processing, vol.2, no.3, 2005, pp. 183-202.

[17] Hazem M. El-Bakry, "New High Speed Normalized Neural Networks for Fast Pattern Discovery on Web Pages," International Journal of Computer Science and Network Security, vol.6, No. 2A, February 2006, pp.142-152.

[18] Hazem M. El-Bakry "Fast Iris Detection for Personal Verification Using Modular Neural Networks," Lecture Notes in Computer Science, Springer, vol. 2206, October 2001, pp. 269-283.

[19] Hazem M. El-Bakry, and Qiangfu Zhao, "Fast Normalized Neural Processors For Pattern Detection Based on Cross Correlation Implemented in the Frequency Domain," Journal of Research and Practice in Information Technology, Vol. 38, No.2, May 2006, pp. 151-170.

[20] Hazem M. El-Bakry, and Qiangfu Zhao, "High speed time delay Neural Networks," International Journal of Neural Systems, vol. 15, no.6, December 2005, pp.445-455.

[21] Hazem M. El-Bakry, and Qiangfu Zhao, "Speeding-up Normalized Neural Networks For Face/Object Detection," Machine Graphics & Vision Journal (MG&V), vol. 14, No.1, 2005, pp. 29-59.

[22] Hazem M. El-Bakry, and Qiangfu Zhao, "A New Technique for Fast Pattern Recognition Using Normalized Neural Networks," WSEAS Transactions on Information Science and Applications, issue 11, vol. 2, November 2005, pp. 1816-1835.

[23] Hazem M. El-Bakry, and Qiangfu Zhao, "Fast Complex Valued Time Delay Neural Networks," International Journal of Computational Intelligence, vol.2, no.1, pp. 16-26, 2005.

[24] Hazem M. El-Bakry, and Qiangfu Zhao, "Fast Pattern Detection Using Neural Networks Realized in Frequency Domain," Enformatika Transactions on Engineering, Computing, and Technology, February 25-27, 2005, pp. 89-92.

[25] Hazem M. El-Bakry, and Qiangfu Zhao, "Sub-Image Detection Using Fast Neural Processors and Image Decomposition," Enformatika Transactions on Engineering, Computing, and Technology, February 25-27, 2005, pp. 85-88.

[26] Hazem M. El-Bakry, and Qiangfu Zhao, "Face Detection Using Fast Neural Processors and Image Decomposition," International Journal of Computational Intelligence, vol.1, no.4, 2004, pp. 313-316.

[27] Hazem M. El-Bakry, and Qiangfu Zhao, "A Fast Neural Algorithm for Serial Code Detection in a Stream of Sequential Data," International Journal of Information Technology, vol.2, no.1, pp. 71-90, 2005.

[28] Hazem M. El-Bakry and Nikos Mastorakis, "Fast Code Detection Using High Speed Time Delay Neural Networks," Lecture Notes in Computer 8Science, Springer, vol. 4493, Part III, May 2007, pp. 764-773.

[29] Hazem M. El-Bakry, and Nikos Mastorakis, "A New Fast Forecasting Technique using High Speed Neural Networks," WSEAS Transactions on Signal Processing, Issue 10, vol. 4, October 2008, pp. 573-595.

[30] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series, " Math. Comput. 19, 1965, pp. 297–301.

[31] R. Klette, and Zamperon, "Handbook of image processing operators," John Wiley & Sonsltd, 1996.

[32] http://www.worsleyschool.net/science/files/virus/page.html

[33] http://en.wikipedia.org/wiki/Virus

[34] http://medical-dictionary.thefreedictionary.com/Biological+virus

[35] http://www.emc.maricopa.edu/faculty/farabee/biobk/biobookdiversity_1.html

[36] http://www.learnartificialneuralnetworks.com/hopfield.html

[37] http://en.wikipedia.org/wiki/Hopfield_net

[38] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities", Proceedings of the National Academy of Sciences of the USA, vol. 79 no. 8 pp. 2554-2558, April 1982.

[39] http://reference.wolfram.com/applications/neuralnetworks/NeuralNetworkTheory/2.7.0.html

[40] http://www.engineeringletters.com/issues_v14/issue_1/EL_14_1_23.pdf

[41] http://www.codeproject.com/KB/recipes/HopfieldNeuralNetwork.aspx

[42] http://web-us.com/brain/neur_hopfield.html

[43] http://www.heatonresearch.com/articles/2/page5.html

[44] Hongmei. He, Ondrej. Sykora, " A Hopfield Neural Network Model for the Outerplanar Drawing Problem," International Journal of Computer Science, vol. 32, no. 4, 2006, available on line, http://www.iaeng.org/IJCS/issues_v32/issue_4/IJCS_32_4_17.pdf

[45] S. Amari, "Learning Patterns and Pattern Sequences by Self-Organizing Nets of Threshold Elements," IEEE Transactions on Computers, vol. C-21, no. 11, pp. 1197–1206, November 1972.

[46] S. Amari and K. Maginu, "Statistical Neurodynamics of Associative Memory," Neural Networks, vol. 1, pp. 63–73, 1988.

[47] D. Hebb, The Organization of Behavior, New York, New York: John Wiley and Sons, 1949.

[48] J. Hopfield, "Neurons with Graded Response Have Collective Computational Properties Like Those of Two-State Neurons," Proceedings of the National Academy of Science USA, vol. 81, pp. 3088–3092, May 1984.

[49] J. Hopfield and D. Tank, "Computing with neural circuits: A model," Science, vol. 233, pp. 625–633, 1986.

[50] I. Arizono, A. Yamamoto, and H. Ohta, "Scheduling for minimizing total actual flow time by neural networks," International Journal of Production Research, vol. 30, no. 3, pp. 503–511, March 1992.

[51] B. Lee and B. Sheu, "Modified Hopfield Neural Networks for Retrieving the Optimal Solution," IEEE Transactions on Neural Networks, vol. 2, no. 1, pp. 137–142, January 1991.

[52] R. Lippmann, "An Introduction to Computing with Neural Nets," IEEE Acoustics, Speech and Signal Processing Magazine, pp. 4–22, April 1987.

[53] M. Lu, Y. Zhan, and G. Mu, "Bipolar Optical Neural Network with Adaptive Threshold," Optik, vol. 91, no. 4, pp. 178–182, 1992.

[54] W. McCulloch and W. Pitts, "A logical calculus of the ideas imminent in nervous activity," Bulletin of Mathematical Biophysics, vol. 5, pp. 115–133, 1943.

[55] R. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," Psychological Review, vol. 65, pp. 386–408, 1958.

[56] R. Rosenblatt, Principles of Neurodynamics, New York, New York: Spartan Books, 1959.

[57] D. Schonfeld, "On the Hysteresis and Robustness of Hopfiled Neural Networks," IEEE Transactions on Circuits and Systems – II : Analog and Digital Signal Processing, vol. 2, pp. 745–748, November 1993.

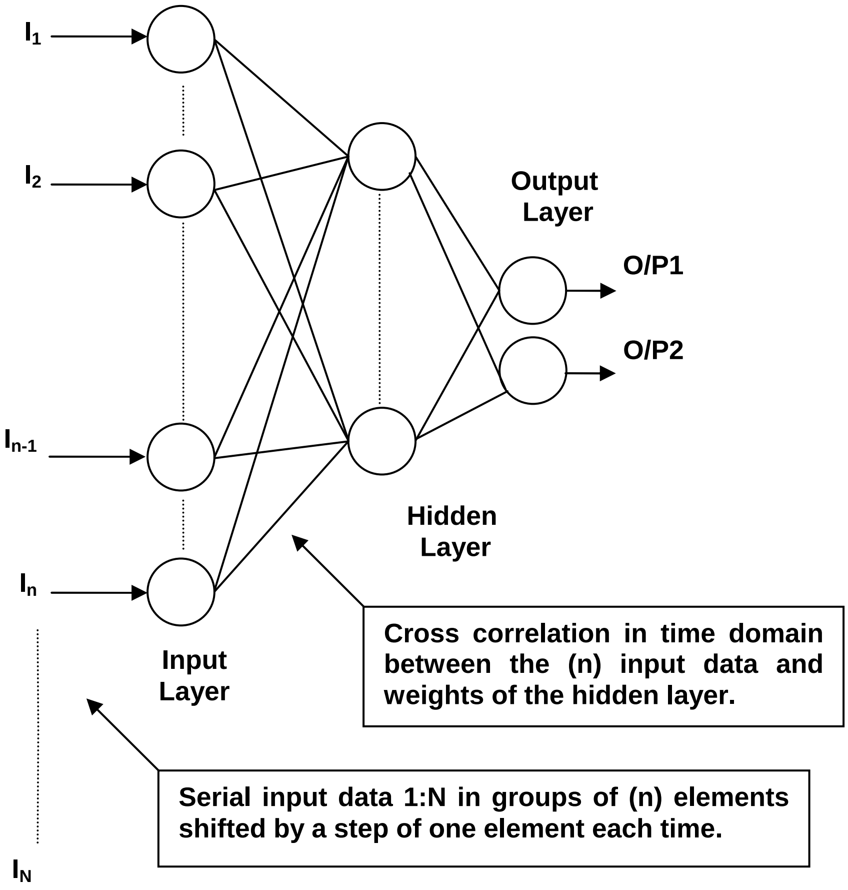

I_1
I_2
I_{n-1}
I_n
I_N
Input Layer
Hidden Layer
Output Layer
O/P1
O/P2
Cross correlation in time domain between the (n) input data and weights of the hidden layer.
Serial input data 1:N in groups of (n) elements shifted by a step of one element each time.

Figure 1. CTDNNs.

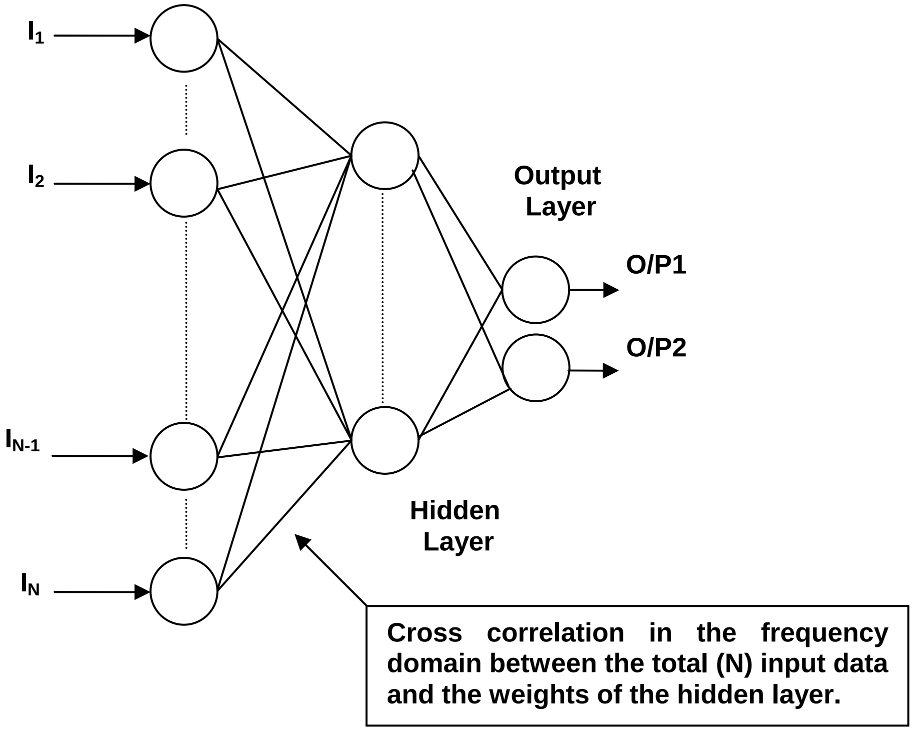

I_1
I_2
I_{N-1}
I_N
Output Layer
O/P1
O/P2
Hidden Layer
Cross correlation in the frequency domain between the total (N) input data and the weights of the hidden layer.

Figure 2. HSTDNNs.

TABLE I: The theoretical speed up ratio for Detecting H1N1 or H1N5 (length of biological virus code=400).

Length of serial data	Number of computation steps required for CTDNNs	Number of computation steps required for HSTDNNs	Speed up ratio
10000	2.3014e+008	4.2926e+007	5.3613
40000	0.9493e+009	1.9614e+008	4.8397
90000	2.1478e+009	4.7344e+008	4.5365
160000	3.8257e+009	8.8219e+008	4.3366
250000	5.9830e+009	1.4275e+009	4.1912
360000	8.6195e+009	2.1134e+009	4.0786
490000	1.1735e+010	2.9430e+009	3.9876
640000	1.5331e+010	3.9192e+009	3.9119

TABLE II: The theoretical speed up ratio for Detecting H1N1 or H1N5 (length of biological virus code=625).

Length of serial data	Number of computation steps required for CTDNNs	Number of computation steps required for HSTDNNs	Speed up ratio
10000	3.5132e+008	4.2919e+007	8.1857
40000	1.4754e+009	1.9613e+008	7.5226
90000	3.3489e+009	4.7343e+008	7.0737
160000	0.5972e+010	8.8218e+008	6.7694
250000	0.9344e+010	1.4275e+009	6.5458
360000	1.3466e+010	2.1134e+009	6.3717
490000	1.8337e+010	2.9430e+009	6.2306
640000	2.3958e+010	3.9192e+009	6.1129

TABLE III: The theoretical speed up ratio for Detecting H1N1 or H1N5 (length of biological virus code=900).

Length of serial data	Number of computation steps required for CTDNNs	Number of computation steps required for HSTDNNs	Speed up ratio
10000	4.9115e+008	4.2911e+007	11.4467
40000	2.1103e+009	1.9612e+008	10.7600
90000	4.8088e+009	4.7343e+008	10.1575
160000	0.8587e+010	8.8217e+008	9.7336
250000	1.3444e+010	1.4275e+009	9.4178
360000	1.9381e+010	2.1134e+009	9.1705
490000	2.6397e+010	2.9430e+009	8.9693
640000	3.4493e+010	3.9192e+009	8.8009

TABLE IV: Practical speed up ratio for Detecting H1N1 or H1N5.

Length of serial data	Speed up ratio (n=400)	Speed up ratio (n=625)	Speed up ratio (n=900)
10000	8.94	12.97	17.61
40000	8.60	12.56	17.22
90000	8.33	12.28	16.80
160000	8.07	12.07	16.53
250000	7.95	17.92	16.30
360000	7.79	11.62	16.14
490000	7.64	11.44	16.00
640000	7.04	11.27	15.89

Enhancement Technique for Leaf Images

N.Valliammal
Assistant Professor, Department of Computer Science,
Avinashilingam Institute for Home Science and Higher
Education for Women, Coimbatore-641 043. INDIA
valli.p.2008@gmail.com

Dr.S.N.Geethalakshmi
Associate Professor, Department of Computer Science,
Avinashilingam Institute for Home Science and Higher
Education for Women, Coimbatore-641 043. INDIA
sngeethalakshmi@yahoo.com

Abstract— **Computer aided identification of plants is an area of research that has gained more attention in recent years and is proving to be a very important tool in many areas including agriculture, forestry and pharmacological science. In addition, with the deterioration of environments, many of the rare plants have died out, and so, the investigation of plant recognition can contribute to environmental protection. A general process of a Computer Aided Plant Classification through Leaf Recognition (CAP-LR) contains four steps, namely, building the leaf database, preprocessing, feature extraction and classification. This paper focuses on the preprocessing step of CAP-LR. In this paper, an approach that simultaneously removes noise, adjusts contrast and enhances boundaries is presented. Experimental results prove that the proposed method is an improved version to the traditional enhancement algorithms.**

Keywords: Contrast Adjustment; Discrete Wavelet Transform; Boundary Enhancement; Median filter.

I. Introduction

Plants are living organisms belonging to the vegetable kingdom that can live on land or in water. They are responsible for the presence of oxygen [1], which is vital for human beings. The ability to know or identify plants allows to assess many important rangeland and pasture variables that are crucial to proper management of plant life. To help botanists in this challenging venture, several researches (Man *et al.*, 2008; Lee and Chen, 2006) are conducted to automatically classify a given input plant into a category. A general process of a Computer Aided Plant Classification Through Leaf Recognition (CAP-LR) contains four steps [5],[6], namely (i) Acquisition of leaf images and creation of plant and leaf image database (ii) Preprocessing the acquired images (iii) Extract salient features and (iv) Cross examine these extracted features with the historical data to match the leaf with its associated plant. The plant that has the maximum match is the recognized plant

Out of these four steps, this paper focuses on the preprocessing stage of CAP-LR. Preprocessing is the technique of enhancing a leaf image in such a way that it increases the efficiency of the subsequent tasks of the leaf recognition system. Leaf images are normally degraded by the presence of noise and low or high contrast both in edge area and image area. Preprocessing an image include removal of noise, edge or boundary enhancement, automatic edge detection, automatic contrast adjustment and segmentation. For segmentation and classification processes [7, 8] CAP-LR is used. In this paper, an approach that simultaneously removes noise, adjusts contrast and enhances boundaries is presented.

II. Proposed Methodology

The proposed algorithm presented presents a novel amalgamation of the existing systems to increase the quality of the image. The method combines the use of CLAHE (Contrast Limited Adaptive Histogram Equalization) algorithm for enhancing the contrast of the input leaf image, Discrete Wavelet Transform (DWT) [2] to identify the edge and non-edge region of the image, edge enhancement using sigmoid function and noise removal using median filter. The various steps involved are shown in Figure 1.

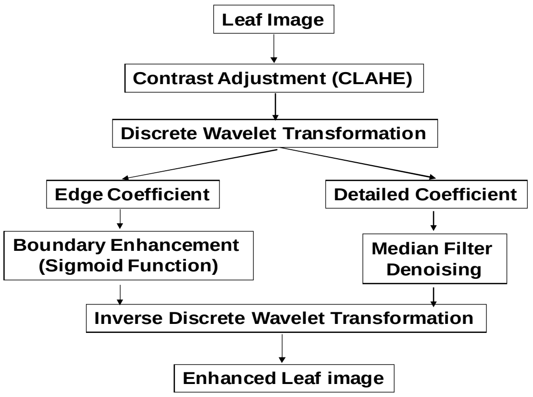

Leaf Image
Contrast Adjustment (CLAHE)
Discrete Wavelet Transformation
Edge Coefficient
Detailed Coefficient
Boundary Enhancement (Sigmoid Function)
Median Filter Denoising
Inverse Discrete Wavelet Transformation
Enhanced Leaf image

Figure 1. Enhancement Procedure

The algorithm begins by applying CLAHE to adjust the contrast of the leaf image. 01CLAHE (Wanga *et al.*, 2004) is a special case of the histogram equalization technique (Gonzalez and Woods, 2007), which seeks to reduce the noise and edge-shadowing effect produced in homogeneous areas. The algorithm is given in Figure 2. In the experiments, NB was set to 64, CL was set to 0.01, tile size used was 8 x 8, and the histogram distribution is Bell-Shaped. The contrast adjusted image is then decomposed using 2D Haar wavelet transform to obtain LL, LH, HL and HH subbands. It is known that the LL subband has the average details of the image, while LH contains horizontal edge details, HL has vertical edge details and HH subband elements contain diagonal edge details. Thus the detailed coefficients are selected. The edge enhancement procedure starts by dividing the wavelet coefficients into 8 x 8 blocks. The image features mean, variance and correlation are calculated for each block to obtain the local information in terms of texture pattern. Figure 2 shows the CLAHE algorithm.

```
Input: Leaf Image, No. of bins (NB),
Clip Limit (CL);

    Output: Contrast Adjusted Image

1. Divide input image into 'n' number of non-
   overlapping contextual regions (tile) of equal
   sizes (8 x 8 used in experiments).
2. For each region
   a. Calculate histogram for each tile using NB
   b. Clip the histogram such that its height does
      not exceed CL (Histogram Redistribution)
      (CL = 0.01 set in experiments).
   c. Use transformation function (Equation ---)
      to create a mapping for this region
Combine neigh-bouring tiles using bilinear
interpolation and modify gray scale values
according to the modified histograms
```

Figure 2. CLAHE Algorithm

Using this information the edges are categorized as strong and weak edges. The weak edges are then enhanced using a sigmoid function (Equation 1).

$$y(x) = \frac{M}{1+e^{-\left(\frac{x-m-\Delta x}{a}\right)}} + \Delta x \qquad (1)$$

where M is 255, m = 128 (for 8 bit image), x is the edge pixel, $-127\ \Delta \le x \le +128$, parameter 'a' refers to the speed of the change around the center.

The next step is to remove the noise from detailed coefficients. For this purpose, a relaxed median filter is used. Traditional median filter is efficient in noise removal. However, the filter sometimes removes sharp corners and thin lines and destroys structural and spatial neighbourhood information. To solve this, this work uses a relaxed median filter (Hamsa *et al.*, 1999). During experimentation, the lower limit was set to 3 and upper limit was set to 5 and the window size used as 3 x 3. After enhancing the edges and removing the noise, finally an inverse wavelet transformation is performed to obtain an enhanced leaf image.

III. Experimental results

The performance metrics used as Peak Signal to Noise Ratio (PSNR), Pratt's Figure Of Merit (FOM) and enhancement speed. All the experiments were conducted in a Pentium IV machine with 2GB Memory and the proposed enhancement algorithm was developed using MATLAB 2009a. The proposed method was evaluated using several test images, three of which is shown as sample in Figure 3 (Leaf1-Leaf6). The manually corrupted images are also shown in Figure 3 (LeafN1-LeafN6)[3]. Fifty percent contrast was added with 10% uniform impulse noise. The results are compared with the traditional median filter and wavelet denoising filter [9]. To compute PSNR, the block first calculates the Mean-Squared Error (MSE) and then the PSNR (Equation 2).

$$\text{PSNR} = 10 \log 10 \left[\frac{R^2}{\text{MSE}} \right] \qquad (2)$$

where MSE = $\frac{\sum_{M,N}[I_1(m,n)-I_2(m,n)]^2}{M*N}$ and R(=255) is the maximum fluctuation in the input image data type, M and N, m and n in MSE equation are number of rows and columns in the input and output image respectively

To compare edge preservation performances of different speckle reduction schemes, the Pratt's figure of merit (Yu and Acton, 2002) is adopted and is defined by Equation (3).

$$\text{FOM} = \frac{1}{\max\{\hat{N}, N_{ideal}\}} \sum_{i=1}^{\hat{N}} \frac{1}{1+d_i^2 \alpha} \qquad (3)$$

where $\hat{N}$ and N_{ideal} are the number of detected and ideal edge pixels, respectively, d_i is the Euclidean distance between the ith detected edge pixel and the nearest ideal edge pixel, and α is a constant typically set to 1/9. FOM ranges between 0 and 1, with unity for ideal edge detection.

Enhancement time is the execution taken by the proposed algorithm to perform the enhancement operation on the noisy image and obtain the reconstructed image. The time is measured in seconds. All the experiments were conducted in a Pentium IV machine with 2GB Memory and the proposed enhancement algorithm was developed using MATLAB 2009a. The proposed method was evaluated using several test image, four of which is shown as sample in Figure 3. The manually corrupted images are shown in Figure 4. Fifty percent contrast was added with 10% uniform impulse noise. The results are compared with the traditional median filter and wavelet denoising filter.

The PSNR and Pratt's Figure of Merit (FOM) values obtained are projected in Table 1. Figure 3 shows the original and corrupted images.

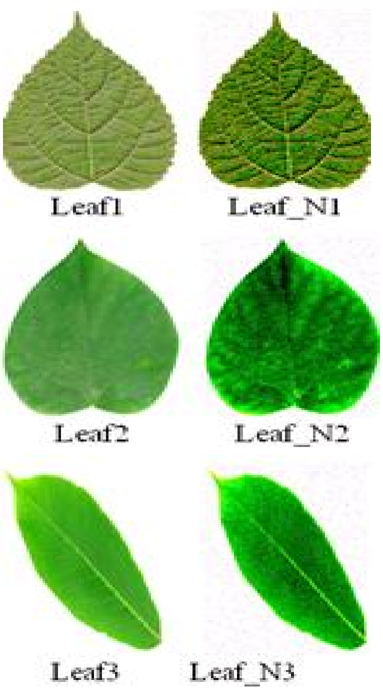


Figure 3. Original and Corrupted Images

TABLE 1. PSNR

Filter Model	Leaf_N1	Leaf_N2	Leaf_N3	Leaf_N4	Leaf_N5	Leaf_N6
Median	31	34	33	34	30	30
Wavelet	34	36	39	40	35	31
Proposed Method	42	46	44	45	41	40

The table 1 & 2 shows the PSNR and FOM value for the different method.

TABLE 2. FOM

Filter Model	Leaf_N1	Leaf_N2	Leaf_N3	Leaf_N4	Leaf_N5	Leaf_N6
Median	0.4004	0.3027	0.4212	0.3072	0.4120	0.4099
Wavelet	0.7399	0.6958	0.7199	0.6841	0.7001	0.7200
Proposed Method	0.7990	0.7437	0.7877	0.7892	0.7813	0.7363

The high PSNR obtained gives the understanding that the visual quality of the denoised image is good. On average the median filter [11] produced an PSNR value of 32 dB, Wavelet produced 35.83dB and 43dB by proposed algorithm. This shows that the proposed method is an improved version of the traditional algorithms. Similarly, while considering the FOM, by the nearing value to unity achieved by the proposed model, it is clear that the proposed model is successful in removing maximum noise [12] from the corrupted image. To compare each filter's performance with respect to FOM performance metric, the average value of the six images were calculated. The median filter based enhancement algorithm showed 0.38, wavelet showed 0.71 and proposed method showed 0.77 FOM. This shows that the proposed algorithm produces better FOM than all the other models indicating that the edge preserving capability is high.

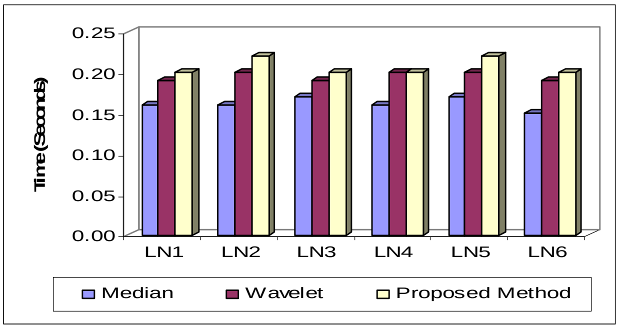


Figure 4. Enhancement Speed

The above figure shows the enhancement speed. While considering the execution time, the median filter was the quickest in enhancing the corrupted image, which was followed by wavelet. The proposed algorithm was the slowest of all the three algorithms. The reason might be because of the extra computations performed by the CLAHE algorithm. However, this difference is very small (0.05 and 0.01 seconds with median and wavelet filters respectively) and can be considered negligible. From the results, it is evident that the speed of the proposed denoising algorithms is faster and the PSNR value obtained is also high.

IV CONCLUSION

Leaf image enhancement is a vital preprocessing step in CAP-LR system. This paper introduced an automatic contrast adjustment, edge enhancement and noise removal algorithm. The algorithm used CLAHE, relaxed median filter and sigmoid function during the enhancement task. The experimental results shows that the proposed method shows significant improvement in terms of noise removed, edge preservation and speed [11]. In future, the impact of the enhancement algorithm on leaf recognition for plant identification [4] is to be studied. Further, methods to automatically calculate the value of NB and CL in CLAHE will also be considered.

REFERENCES

[1] Palmer, J.D., Adams, K.L., Cho, Y., Parkinson, C.L., Qiu, Y.L. and Song, K. (2000) Dynamic Evolution of Plant Mitochondrial Genomes: Mobile Genes and Introns and Highly Variable Mutation Rates, Proceedings of the National Academy of Sciences of the United States of America, Vol. 97, No.13, Pp. 6960-6966.

[2] Gu, X., du, J. and Wang, X. (2005) Leaf Recognition Based on the Combination of Wavelet Transform and Gaussian Interpolation , Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Vol. 3644/2005, Pp. 253-262.

[3] Sathyabama, B., Mohanavalli, S., Raju, S. and Ahbaikumar, V. ((2011) Content Based Leaf Image Retrieval (CBLIR) Using Shape, Color and Texture Features, Indian Journal of Computer Science and Engineering (IJCSE), Vol 2, No. 2, Pp. 202-211.

[4] Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y., Chang, Y. and Xiang, Q. (2007) A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network, IEEE International Symposium on Signal Processing and Information Technology, Pp. 11-16.

[5] N. Valliammal , S.N.Geethalakshmi, Analysis of the Classification Techniques for Plant Identification through Leaf Recognition, CIIT International Journal of Data Mining Knowledge Engineering, Vol.1, No.5, August 2009.

[6] N.Valliammal, S.N.Geethalakshmi, Leaf Recognition for Plant Classification, IETECH, International Engineering and Technology Journal of Advanced Computations, Vol.3, No.3, 2009.

[7] N. Valliammal, S.N.Geethalakshmi, Performance Analysis of Various Leaf Boundary Edge Detection Algorithms, A2CWic'10, Proceedings of the First ACM-W celebration of Women in Computing in India,16-17, September 2010.

[8] N. Valliammal, S.N.Geethalakshmi, Hybrid Image Segmentation Algorithm for Leaf Recognition and Characterization, International Conference on Process Automation, Control and Computing, PACC 2011,20-22 July 2011.

[9] Li, Y., Zhang, Y., Zhu, J. and Li, L. (2010) Wavelet-based maize leaf image denoising method, World Automation Congress (WAC), Pp. 391-395.

[10] Ma, L., Fang, J., Chen, Y. and Gong, S. (2010) Color Analysis of Leaf Images of Deficiencies and Excess Nitrogen Content in Soybean Leaves, International Conference on on E-Product E-Service and E-Entertainment (ICEEE), Pp.1-3.

[11] El-Helly, M., Rafea, A. and El-Gammal, S. (2003) An integrated image processing system for leaf disease detection and diagnosis, 1st Indian International Conference on AI (IICAI-0), Hyderabad, India.

Zhang, J. (2010) An efficient median filter based method for removing random-valued impulse noise, Digital Signal Processing, Vol. 20, Issue 4, Pp. 1010-1018.

[12] Rubio, E.L. (2010) Restoration of images corrupted by Gaussian and uniform impulsive noise, Pattern Recognition, Vol.43, No.5, Pp. 1835-1846

Secret Sharing Scheme based on Chinese reminder theorem and polynomials interpolation

Qassim AL Mahmoud
Faculty of Mathematics and Computer Science
The University of Bucharest, Romania
qassim_oudat@yahoo.com

***Abstract*: The concept for a secret sharing is necessary to build a security system that saves and retrieves information to avoid its loss or theft, and increase the security infrastructure. Secret sharing schemes can be used for any way in which the access to an important resource has to be restricted. consideration to the concept of secret must be taken into account the group of people selected to be the group authorized to build the concept of secret sharing, dividing this group into subsets where each subset can retrieve private confidence. this paper build scheme combine from Chinese reminder theorem and interpolation polynomials which depend on the tow famous thresholds secret sharing scheme, Mignotte' Scheme, and Shamir scheme respectively in order to produce flexible and extensible frame work for secret sharing.**

Keywords*: secret sharing scheme, threshold secret sharing scheme, Shamir secret sharing, Mignotte secret sharing.

I. Introduction

The most important properties of secret sharing is that it is secret, to preserve the secret from being lost or stolen, as well as building a system that is not based on dictatorship (i.e. rely only on one person who owns a secret to access the information stored in the database).

From this point, the need of a concept for a secret sharing is necessary to build a security system that saves and retrieves information to avoid its loss or theft, and increase the security infrastructure. So we have all the security status of access ways. To illustrate this, let us consider the banking system as a simple example where it is necessary to secure (save and store) customers' information from the staff themselves. The problem is that allowing employees to access such information to make a modifications requires to know the secret, but in the same time that secret cannot be given to all staff in the bank. In addition, given the secret to the bank's manager is not practical as his presence is not always necessary to grant the employees access at any moment needed. Even though the president's presence always makes an effective and safe way to access information(because the occurrence of any urgent matter), the president may however loss the secret which will cause to a loss of a important information. To prevent information lost, it is necessary to think of a more secure access to information without relying only on a single person.

What is the concept of secret sharing? And how can we get this application secured? Many questions will be answered by this research paper.

Secret sharing schemes have been introduced by Blakley [1] and Shamir [2] independently as a solution for safeguarding cryptographic keys. Secret sharing schemes can be used for any way in which the access to an important resource has to be restricted. consideration to the concept of secret must be taken into account the group of people selected to be the group authorized to build the concept of secret sharing, dividing this group into subsets where each subset can retrieve private confidence. In fact this is the definition of access structure. In this research the mathematically concept of access structure will be mention. In order to understand the secret sharing. Let us look at the secret, we can derive information; called shares or shadows ; that are authorized to distribute to the group so that only a fixed number(t) of people (or more) may restore that secret. Less than satisfy number of people(t-1) should not be able to know anything about that secret, this way is called threshold secret sharing scheme.

Secret sharing has tow algorithms, first is shares generation algorithm that distributes the shares of participants, and the second is reconstruction algorithm for secret.

The most important two schemes that depend on the threshold scheme(Shamir secret sharing scheme and Mignotte's scheme). Shamir scheme generation algorithm is based on polynomials in order to distribute shares of participants, and reconstruction algorithm is based on polynomials interpolation. The Mignotte's threshold secret sharing scheme is based on the Chinese reminder theorem both generation and construction algorithms with special properties of prim numbers in number theory. Through our understanding of these two schemes, we can present our approach is evident in this research. We will then see how our scheme can generate the shares in generation algorithm for all participants based on the Chinese reminder theorem in order to distribute the shares and recover the secret in reconstruction algorithm depends on the polynomials interpolation.

In the rest of this chapter we will mention the concept of Access structure and some of the basic theorem of Chinese reminder theorem. The second chapter it will be the previous studies which is divided into two studies, the first study will be explained to a threshold Shamir secret sharing scheme. The second study will be an explanation of the Mignotte's threshold secret sharing scheme. In the third chapter we will offer a presentation to our scheme with illustration by an example of small artificially. In Chapter four it will be the conclusion for our scheme .

A. Access structure

Let $X = \{1, 2, \ldots, \mathrm{n}\}$ be the set of users, The access structure $\Gamma \subseteq \mathrm{P}(X)$ is the set of all qualified subsets. We give bounds on the amount of information(shares) for each participant. Then we apply this to construct computational schemes for general access structures. The size of shares each participant must have in our schemes is nearly minimal for $\{1, 2, \ldots, \mathrm{n}\}$ let us consider a set of groups $\Gamma \subseteq \mathrm{P}(X)$ The (authorized) Access structure of a secret sharing scheme is the set of all groups which are designed to reconstruct the secret. The elements of the access structure A will be referred to as the authorized groups/sets and the rest are called unauthorized groups/sets.

Saito, and Nishizeki have remarked [3] any access structure must satisfy the natural condition (i.e. that if a group can recover the secret, so can a larger group). Benaloh and Leichter [4] called such access structures monotone .

The unauthorized access structure $\overline{\Gamma}$ is well specified by the set of the maximal unauthorized groups.

In the secret sharing schemes the number of the participants in the reconstruction phase was important for recovering the secret. Such schemes have been referred to as" threshold secret sharing schemes."

Definition 1: Let $n \geq 2,\ 2 \leq k \leq n$. The access structure $\Gamma = \{A \in P(\{1,2,\ \ldots,\ n\} / |\mathrm{A}| \geq \mathrm{k})\ \}$ will be referred to as the $(k,\ n)$ -threshold access structure.

In case $\Gamma = \{1, 2,\ \ldots,\ n\}$, an Γ -secret sharing scheme will be referred to as a unanimous consent secret sharing scheme of rank n . In these schemes, the presence of all users is required in order to recover the secret. A unanimous consent secret sharing scheme of rank n is equivalent with an $(n,\ n)$ -threshold secret sharing scheme and, thus, any $(n,\ n)$ -threshold secret sharing scheme can be used in order to realize unanimous consent, for more details the reader have to read in [5], [6].

B. Chinese Reminder Theorem (CRT)

The Chinese Remainder Theorem gives solutions to systems of congruencies with relatively prime moduli. The solution to a system of congruence with relatively prime moduli may be produced using a formula. by computing modular inverses, or using an iterative procedure involving successive substitution.

The Chinese Remainder Theorem says that certain systems of simultaneous congruencies with different moduli have solutions. The idea embodied in the theorem was apparently known to Chinese mathematicians a long time ago — hence the name.

We will begin by collecting some useful lemmas without prove to help us understanding (CRT)[7].

Lemma 1. Let m and $\mathrm{a}_1, \ldots, \mathrm{a}_n$ be positive integers. If m is relatively prime to each of $\mathrm{a}_1, \ldots, \mathrm{a}_n$, then it is relatively prime to their product $\mathrm{a}_1 \ldots \mathrm{a}_n$

We call the greatest common divisor (a, b) of a and b is greatest in the sense that it is divisible by any common divisor of a and b. The next result is the analogous statement for least common multiples.

Lemma 2. Let m and $\mathrm{a}_1, \ldots, \mathrm{a}_n$ be positive integers. If m is a multiple of each of $\mathrm{a}_1, \ldots, \mathrm{a}_n$, then m is a multiple of $[\mathrm{a}_1, \ldots, \mathrm{a}_n]$.

Lemma 3. Let $\mathrm{a}_1, \ldots, \mathrm{a}_n$ be positive integers. If $\mathrm{a}_1, \ldots, \mathrm{a}_n$ are pairwise relatively prime (that is $(\mathrm{a}_i,\ \mathrm{a}_j) = 1 \text{ for } \mathrm{i} \neq \mathrm{j}$), then $[\mathrm{a}_1, \ldots, \mathrm{a}_n] = \mathrm{a}_1 \ldots \mathrm{a}_n$.

Theorem 1. (The Chinese Remainder Theorem(CRT)): Suppose $p_1, \ldots, p_n$ are pairwise relatively prime (that is, $(p_i, p_j) = 1 \text{ for } \mathrm{i} \neq \mathrm{j}$). Then the system of congruence :

$$\begin{aligned} x &= a_1 \pmod{p_1} \\ x &= a_2 \pmod{p_2} \\ &. \\ &. \\ &. \\ x &= a_n \pmod{p_n} \end{aligned}$$

has a unique solution $\mathrm{mod}(p_1 \cdot \cdot \cdot p_n)$.

II. Previous Study

The previous studies which is divided into two sections, the first section will be explained to a threshold Shamir secret sharing scheme[8] based on polynomials interpolation.

The second section will be an explanation of the Mignotte's threshold secret sharing scheme based on (CRT)[9].

A. Threshold Shamir Secret Sharing Scheme

In this section, we first review Shamir's threshold secret sharing scheme. Then we will mention some important definitions about Shamir secret sharing scheme.
In Shamir's (t; n) scheme based on Lagrange interpolating polynomial, there are n shareholders, $P = \{P_1,..., P_n\}$, and a dealer D. The scheme consists of two algorithms:

generation Shares algorithm: dealer D first picks a polynomial f(x) of degree (t-1) randomly such $f(x) = a_0 + a_1 x + ... + a_{t-1}$, in which the secret s = a_0 and all coefficients $a_0, a_1, ..., a_{t-1}$ are in a finite field Fp = GF(p) with p elements, where s < p, and D computes: $s_1 = f(1), s_2 = f(2), ... s_n = f(n)$
Then, D outputs a list of n shares, $(s_1, s_2, ..., s_n)$, and distributes each share to corresponding shareholder privately.

Secret reconstruction algorithm: with any t shares, $(s_{i1}, s_{i2}, ..., s_{it})$
where $A = \{i_1, ..., i_t\} \subseteq \{1, 2, ..., n\}$.

We can reconstruct the secret s as follows.

$$s = f(0) = \sum_{i \in A} s_i \left(\prod_{j \in A - \{i\}} \frac{x_j}{x_j - x_i} \right)$$

We note that the above scheme satisfies basic requirements of secret sharing
scheme as follows:

1) with knowledge of any t or more than t shares, it can reconstruct the secret s.
2) with knowledge of any fewer than t shares, it cannot reconstruct the secret s.

Shamir's scheme is information-theoretically secure since the scheme satisfies these two requirements without making any computational assumption. For more information on this scheme, readers can refer to the original paper [10].

Definition 2 (Information rate). Information rate of a secret sharing scheme is the ratio between the length, in bits, of the secret and the maximal length of shares distributed to shareholders. Let a be the number of bits of the secret and $b = \max_{i \in \{1,...n\}} \{b_i\}$ be the number of bits of maximal share. The information rate is defined as.

$$\rho = \frac{a}{b}$$

The secret sharing scheme is ideal if $\rho = 1$.

Definition 3 (Perfect threshold secret sharing [11]). We say that a (t, n) threshold secret sharing scheme is perfect if any (t-1) or fewer than (t-1) shareholders who work together with their corresponding shares cannot get any information, in the information-theoretic sense, about the secret.

Shamir's secret sharing scheme is perfect. If we use entropy to describe this perfect secret property of threshold secret sharing scheme, Karnin et al. [12] have shown that in all perfect schemes, the length of share must be larger than or equal to the length of the secret s. In other words, the information rate of all perfect schemes is no more than 1.

B. Mignotte's Threshold Secret Sharing Scheme

Mignotte's Scheme is the most important threshold secret sharing schemes based on the Chinese remainder theorem.

In [13] uses special sequences of integers, referred to as Mignotte sequences.

Definition 4. Let n be an integer, n ≥ 2, and 2 ≤ k ≤ n. An (k, n)- Mignotte sequence is a sequence of pairwise coprime positive integers $p_1 < p_2 < \cdot \cdot < p_n$ such that

$$\prod_{i=0}^{k-2} P_{n-i} < \prod_{i=1}^{k} P_i .$$

Given a publicly known (k, n)-Mignotte sequence, the scheme works as follows:
• The secret S is chosen as a random integer such that β < S < α, where $\alpha = \prod_{i=1}^{k} P_i$ and $\beta = \prod_{i=0}^{k-2} P_{n-i}$;
• The shares I_i are chosen as $I_i = S \bmod p_i$, for all 1 ≤ i ≤ n;
• Given k distinct shares $I_{i1}, \ldots, I_{ik}$, the secret S is recovered using the standard Chinese remainder theorem, as the unique solution modulo $P_{i1} \cdot \cdot P_{ik}$ of the system :

$$x \equiv I_{i1} \bmod P_{i1}$$

$$\cdot$$

$$\cdot$$

$$\cdot$$

$$x \equiv I_{ik} \bmod P_{ik}$$

Indeed, the secret S is an integer solution of the above system by the choice of the shadows. Moreover, S lies in $Z_{Pi1....Pik}$,because S < α. On the other hand, having only k −1 distinct shares $I_{i1}, \ldots, I_{ik}$, we obtain only that $S \equiv x_0 \bmod P_{i1} \cdot \cdot P_{ik}$, where x_0 the unique solution.

In order to assure a level of security, (k, n)-Mignotte sequences with a large factor must be chosen.
Iftene in [14] extended the (k, n)-Mignotte to be generalized (k, n)-Mignotte , in other word we can apply this scheme not only on coprime numbers, he extended for any positive integer numbers.

III. Secret Sharing Scheme based on (CRT) and polynomials interpolation

Let $P = \{p_1, \dots p_n\}$ be a set of pairwise prime numbers and let $\{a_1, \dots a_n\}$ be a set of integers such that the system of congruence of Chinese reminder theorem given by :

$$\begin{aligned} x &\equiv a_1 \bmod p_1 \\ x &\equiv a_2 \bmod p_2 \\ &. \\ &. \\ x &\equiv a_n \bmod p_n \end{aligned}$$

This is system of equations has a unique solution in $Z_{p_1 p_2 \cdots p_n}$. This is mean there exist one and only one solution such that this solution bounded $(x < \prod_{i=1}^{n} p_i$) .
Now .
There exist such integers $\{q_1, \dots q_n\}$ corresponding to $\{p_1, \dots p_n\}$ and $\{a_1, \dots a_n\}$, respectively.
where :

$$\begin{aligned} (x - a_1) &= p_1 q_1 \\ (x - a_2) &= p_2 q_2 \\ &. \\ &. \\ (x - a_n) &= p_n q_n \end{aligned}$$

Where $p_i q_i$ secret for all $i = 1$ to n .
We can construct as equation of degree (n) from up system as with (n) of solutions one and only one of these solutions $x \in Z_{p_1 p_2 \cdots p_n}$ the form of this equation as :

$$(x - a_1)(x - a_2)\dots(x - a_n) = (p_1 p_2 \dots p_n)(q_1 q_2 \dots q_n).$$

Imply the equation of degree (n) as :

$$x^n - C_1 x^{n-1} + C_2 x^{n-2} - C_3 x^{n-3} + \dots \pm C_m x^{n-m} \pm C_{m+1} x^{n-(m+1)} \pm \dots \pm C_n = (\prod_{i=1}^{n} p_i q_i) \quad (1)$$

Where the sign (±) for the coefficients C_m take as follow :

$$C_m = \begin{cases} +C_m & if \quad n \text{ is odd} \wedge m \text{ is even} \\ -C_m & \text{if} \quad n \text{ is even} \wedge m \text{ is odd} \end{cases}$$

And C_1 , C_2 , . . . , C_n take values as follow :
Now we will construct our scheme as follow :
Before start construct the algorithms for scheme we have to define some sets important to understand our scheme.
Let $N = \{1, 2, \dots, n\}$ a set of users and let $P = \{p_1, \dots p_n\}$ a set of a pairwise prime number defined in up, and we define B as the set of all sets of size (k), the number of primes in the set .

$$B = \left\{ \{p_1, \dots p_n\}^k / \forall A, B \in \{p_1, \dots p_n\}^k, (\exists p_i \in A \wedge p_i \notin B), \forall i \in N, \ 2 \le k \le n \right\}$$

.

This is mean $|B| = \binom{n}{k} = \dfrac{n!}{k!(n-k)!}$

We will define the secret space X as :

$$\text{X} = \left\{ \text{x} / \text{ x integer} \wedge \forall \text{A} \in B \ / \ (\text{ x} < \min \left\{ \prod_{p_i \in A} p_i \right\}) \right\}$$

We also define the set $C \subseteq B$ is the set of all sets satisfy the condition in the secret space X as :

$$\text{C} = \left\{ A \ / \ (\text{ x} < \left\{ \prod_{p_i \in A} p_i \right\}) \ , A \in B \right\}.$$

For the secret chosen x from the secret space X .
Now
The generation shares algorithm: work as follow :.
any users $i \in N$ has a set of possible shares $\left\{ (a_i , \prod_{p_i \in A} p_i q_i) \ / \ \forall \ A \in B \right\}$,q_i corresponding for a_i , p_i respectively, such that

$$x \equiv a_i \mod p_i \quad \forall \ i = 1, \dots, n$$

We see for any integer prime (p_i) may be belong for some difference sets $A \in B$, this mean $\forall \ i \in N$ users has some shares depend of the position of $p_i \in A$ and $\forall \ A \in B$, Then we have to construct the share space S such as :

$$S = \left\{ (a_i , \prod_{P_i \in A} p_i q_i) / \left(\forall i \in N \right) \wedge \left(\forall p_i \in A \right) \wedge \left(\forall A \in B \right) \right\}$$

$$|S| = n \times |S_i|$$

where $|S_i|$ the number of shares for user *i*

Define as follow:

$$S_i = \left\{ (a_i , \prod_{P_i \in A} p_i q_i) / \ \left(\forall p_i \in A \right) \wedge \left(\forall A \in B \right) \right\} \quad \forall i \in N$$

$|S_i| = [k \times (n-k)]$ $\forall S_i$ the number of shares for every user.

It is important to construct Access structure Γ such as :

$$\Gamma = \left\{ \mathrm{D} / \mathrm{D} \in \{1,...,\mathrm{n}\}^k , \forall i \neq j \in D / (p_i \wedge p_j) \in A, \ A \in B \Leftrightarrow \prod_{P_i \in A} p_i q_i = \prod_{P_j \in A} p_j q_j \right\}$$

The integer k the same integer which we defined in the set (B) in up definition, and called the threshold k , and such this Access structure Γ called (k, n) – Threshold Access structure, and the scheme called (k, n) – threshold secret sharing scheme.

The reconstruct algorithm: any distinguish k of users can construct the secret x by applying the equation (*) using their shares and find the solution x, in equation which construct from (1).

We illustrate the scheme in below example.

Example : (with artificially small parameters) .

Let $N = \{1, 2, 3\}$ set of users and let $P = \{5, 3, 7\}$

Then $n = 3$,

let $k = 2$, then The set

$$B = \{\{5,3\},\{5,7\},\{3,7\}\} \qquad |B| = \frac{n!}{k!(n-k)!} = 3$$

The secret space $X = \{x \ / x < \min\{15, 35, 21\}\}$

$$X = \{x \ / x < 15\}$$

Now let the dealer chose the secret $x = 10$ then he can construct the system of Chinese reminder theorem in order to find $\{a_1, a_2, a_3\}$ and $\{q_1, q_2, q_3\}$ as follow :

$$x \equiv 0 \bmod 5$$
$$x \equiv 1 \bmod 3$$
$$x \equiv 3 \bmod 7$$

Imply :

$$a_1 = 0 \qquad p_1 = 5$$
$$a_2 = 1 \qquad p_2 = 3$$
$$a_3 = 3 \qquad p_3 = 7$$

Then the corresponding $\{q_1, q_2, q_3\}$ for $\{a_1, a_2, a_3\}$ and $\{p_1, p_2, p_3\}$, respectively , it is will be as follow:

$$q_1 = 2$$
$$q_2 = 3$$
$$q_3 = 1$$

Now the dealer construct the share space S as follow :

$$S = \{(0,90),(0,70),(1,90),(1,63),(3,70),(3,63)\}$$

$$|S| = n \times |S_i| = 6$$

The form of shares as point $(a_i , y_j) \ \forall i \in N , \ j = 1 \, to \, |S_i|$.

The dealer distribute the shares for users $N = \{1, 2, 3\}$ as follow :

$$S_1 = \{(0,90),(0,70)\} \qquad |S_1| = [k \times (n-k)] = 2$$
$$S_2 = \{(1,90),(1,63),\} \qquad |S_2| = [k \times (n-k)] = 2$$
$$S_3 = \{(3,70),(3,63)\} \qquad |S_3| = [k \times (n-k)] = 2$$

For users $\{1, 2, 3\}$, respectively.

Any tow users can reconstruct the secret x by pooling their share when the y-axis of their points equal from difference users

(i.e. reconstruct secret x if and only if $(a_i \neq a_j)$ $\wedge$ $(y_i = y_j)$)

Let consider $\{1, 3\}$ users then they have 2 shares with same $y_i = y_j$

The shares from $\{1, 3\}$ can reconstruct the secret x applying the equation (1) by their shares $\{(0,70),(3,70)\}$

Then the users build the equation of degree (2) as what we define in previous :

$$(x - 0)(x - 3) = 70$$
$$x^2 - x = 70$$
$$x^2 - 3x - 70 = 0$$

The solutions for this equation are : $x = 10$ and $x = -7$

Then the secret it will be a unique solution in Z_{15} then $x = 10$.

In this scheme each users has $\left[k \times (n-k)\right]$ shares, the group shares for the same user cant reconstruct the secret alone , He can use one share with different other users with difference their shares to reconstruct the secret. In addition we can in future study develop this scheme to use it in many features of secret sharing (i.e. for example we can release compartments Access structures, or we can use it for verifiable secret sharing scheme, etc).

The security of this scheme depend of the hard of factorization problem, so the chosen large number of shares make the scheme more secure.

IV. Conclusion

The main idea of this paper in order to build scheme combine from Chinese reminder theorem and interpolation polynomials which depend on the tow famous thresholds secret sharing scheme, Mignotte' Scheme, and Shamir scheme respectively. obviously it is secure as long as the hard of factorization problem. So it is computational- secure scheme, for this reason we want in future study for this scheme be more secure.

References

[1] A. Shamir. How to share a secret. Communications of the ACM, 1979.

[2] G. R. Blakley. Safeguarding cryptographic keys. In National Computer Conference, 1979, volume 48 of American Federation of Information Processing Societies Proceedings, pages,1979.

[3] M. Ito, A. Saito, and T. Nishizeki. Secret sharing scheme realizing general access structure. In Proceedings of the IEEE Global Telecommunications Conference, Globecom '87, pages 99–102. IEEE Press, 1987.

[4] J. Benaloh and J. Leichter. Generalized secret sharing and monotone functions. In S. Goldwasser, editor, Advanced in Cryptology-CRYPTO' 88, volume 403 of Lecture Notes in Computer Science, pages 27–35. Springer-Verlag, 1989.

[5] E. D. Karnin, J. W. Greene, and M. E. Hellman. On secret sharing systems. IEEE Transactions on Information Theory, IT-29(1):35–41, 1983.

[6] Sorin Iftene: Secret Sharing Schemes with Applications in Security Protocols. Sci. Ann. Cuza Univ.2-5 (2007).

[7] Johannes A . Buchmann : introduction to cryptography(second edition).51-54, Springer, 2004.

[8] Sorin Iftene: Secret Sharing Schemes with Applications in Security Protocols. Sci. Ann. Cuza Univ.12-14, (2007).

[9] Sorin Iftene: Secret Sharing Schemes with Applications in Security Protocols. Sci. Ann. Cuza Univ.14-16, (2007).

[10] A. Shamir How to share a secret, Communications. ACM, 22(11) (1979), 612- 613.

[11] A. J. Menezes, P. C. Oorschot, S. A. Vanstone, Handbook of applied cryptography. CRC Press, Oct. 1996.

[12] E. D. Karnin, J. W. Greene, M. E. Hellman, On Secret Sharing Systems, IEEE Trans. on Information Theory., 29(1) (1983) 35- 40.

[13] M. Mignotte. How to share a secret. In T. Beth, editor, Cryptography-Proceedings of the Workshop on Cryptography, Burg Feuerstein, 1982, volume 149 of Lecture Notes in Computer Science, pages 371–375. Springer-Verlag, 1983.

[14] Sorin Iftene: Secret Sharing Schemes with Applications in Security Protocols. Sci. Ann. Cuza Univ.15-16, (2007).

ENHANCING COMMUNITY POLICING USING A VIRTUAL COMMUNITY MODEL

Rufai M. M. and Adigun J. O
Dept. of Computer Technology,
Yaba College of Technology, Lagos, Nigeria
Email:m_rufai@yahoo.com

Abstract - **Globalisation and Information Communication Technology have both exposed people to perverted foreign cultures with associated criminal tendencies. Consequently, there has been an increase in the perpetration of crimes in most communities especially in the developing nation like Nigeria. The Nigerian Police has made cogent effort in checking the upsurge of crimes, without significant success. Perhaps, one of the factors responsible for the failure is that the police have not integrated members of the community in their war against crimes or an effective tool has not been employed in reaching members of the community. People have reservations for the Nigerian police on account of three reasons, namely: a) perceived rise in crime/inability of the police to cope with the demand for protection by the citizens, (b) poor perceptions about the ability of the criminal justice system to respond to the needs of the victims of crime and (c) inadequacies of the formal police service. This paper discusses how community policing can be enhanced using virtual community. It describes the modus operandi of existing community policing approach in Nigeria, the associated problems and the changes information technology can make. As part of this research we will review relevant literature on existing virtual communities and we will develop a virtual community model for effective community policing. The paper concludes that community policing can better be enhanced using a virtual community model (VCM).**

Key words: Community Policing, Virtual Community

I. INTRODUCTION

The high rate of crime in urban and rural communities call for a review of our approach to crimes fighting and its prevention. Some of the frequently reported crimes are kidnapping, robbery, murder, terrorism, tribal feud to mention but a few. The survey conducted by Centre for Law Enforcement Education in Nigeria (CLEEN)[1] on crime rates revealed that Murder crime increased from 1629 to 2133 in the year 1994 to 2003. These crimes are perpetrated by members of the community. Some of the factors responsible for the failure of the police in this regard may be lack of understanding between members of the community and the police. A situation where a communication barrier exists between the police and the community residents aggravates the situation. An un-enlightened rural man sees the police as threat to the peace of their land. Their conception is that the police have come to intrude into their privacy or have come to usurp the power of the community head[3]. Consequently, they meet the police with different unwelcome treatment. A community having such wrong impression needs to be enlightened and adequately oriented on the role of the police in combating crime and the need for their support in making the community peaceful.

A. Community Policing Concepts

Community Policing can be defined as a philosophy of or an approach to policing which recognizes the interdependence and shared responsibility of the police and the community in ensuring a safe and secure environment for all the people of the country. Community Policing aims to establish an active and equal partnership between the police and the public through which crime and community safety issues can jointly be determined and solutions designed and implemented. Community policing seeks to achieve the following objectives:

- **Service orientation:** The safety of the community is prioritized. The community is seen as the client and the service need of the client is given proper attention. The service orientation is client-centered.

- **Partnership:** The police see the community as partners in the battle against crime. Consequently, the community needs and policing priorities are determined through consultation with the community.

- **Problem solving:** This relates to the joint identification and analysis of the actual and potential causes of crime and conflict within communities. This analysis guides the development of measures to address such problems over the short-, medium- and long-term.

- **Empowerment:** This refers to the creation of a sense of joint responsibility and a joint capacity for addressing crime, service delivery and safety and security amongst members of the community and The Police Service personnel.

- **Accountability:** Accountability will be realized by creating mechanisms through which the Police can be made answerable for addressing the needs and concerns of the communities they serve.

II THE EXISTING SYSTEM (THE NIGERIA POLICE FORCE)

- **Legal framework for The Nigeria Police Force**

The Nigeria Police Force is constitutionally empowered to provide security for lives and property of Nigerians. This vital security apparatus derives its existence from Section 214 (1) of the 1999 Constitution which stipulates that "there shall be a Police Force for Nigeria, which shall be known as the Nigeria Police Force, and subject to the provisions of this section, no other police force shall be established for the Federation or any part thereof" (The Constitution of Federal Republic of Nigeria. 1999)[9].

Furthermore, Section 4 of the Police Act, 1990 outlines the general duties of the Police as follows : "The police shall be employed for the prevention and detection of crime, the apprehension of offenders, the preservation of law and order, the protection of life and property, and the due enforcement of all laws and regulations with which they are directly charged and shall perform such military duties within or outside Nigeria as may be required of them, by or under the authority of this or any other Act." That these duties of ensuring order, safety and security are important t to the making of a good society is not in doubt.

Section 14 (2) (b) of the 1999 Constitution, stipulates that: "The police shall be employed for the prevention and detection of crime, the apprehension of offenders, the preservation of law and order, the protection of life and property, and the due enforcement of all laws and regulations with which they are directly charged and shall perform such military duties within or outside Nigeria as may be required of them, by or under the authority of this or any other Act."

- **Why People Lost Trust In Traditional Police Structures**

Literature on policing has revealed several reasons why people shun reporting of criminal activities and civil complaints to patronize police authorities; they however resort to informal policing structures. The reasons advanced for not reporting their cases to police include: (a) perceived rise in crime/inability of the police to cope with the demand for protection by the citizens, (b) poor perceptions about the ability of the criminal justice system to respond to the needs of the victims of crime and (c) inadequacies of the formal police service.

Added to the above reasons is the perceived failure of the state to provide citizens with the protection they require[8]. Of the three reasons found in previous studies, the strongest appears to be rise in crime and perceived inadequacies of the police in the provision of safety and security to the citizens, especially the poor.

Jemibewon (2001)[4] also opines that lack of confidence in the police structures appears to be a crucial reason found in the literature on why citizens embrace informal policing structures. The public shun the formal police structure because of community's grievances against the police, these perceived grievances include: corruption, incompetence, brutalisation of citizens and institutional failure.

Furthermore, Del Buono (2003)[2] lends credence to the view above that the police along with the military are among the three most repressive institutions in human society. The police are "largely inactive" in their policing roles, but are active when it comes to harassment of members of the public.

Community Policing Effort in Nigeria Police Force

The concept of community policing in the Nigeria Police Force surfaced when some police officers were sent to England to understudy community policing as practised in the UK. Consequently, in 2004, it was officially launched in six pilot states (i.e. Benue, Enugu, Jigawa, Kano, Ondo and Ogun). In 2008, in line with the president's declaration of 7-point agenda, the then Inspector General of Police introduced Community Policing as both the strategy and philosophy of the entire NPF.

Some of the existing instrument of community policing in Nigeria Police force are:

- **Police/Community Relations Committees (PCRCs) (PCRCs)** is an-ongoing committee setup by the Nigeria Police Force. It works to bring together members of a locality's diverse communities and its police officers to improve

.

community and police relations, further an authentic community policing culture, and promote dignity, understanding, and respect in police and community interaction.

The PCRC has been established in some part of the country to achieve the aforementioned objective. For instance Community Safety Partnerships have been introduced in two Divisions in Lagos and FCT. The senior representatives involved – from Local Government, police, the communities and many other key agencies have made a commitment to work together in the future to gain a full understanding of the local safety issues that affect their communities and work in a partnership to resolve them.

- **Establishment of Community Safety and Security Forums**

Community Safety and Security Forum is one of the recent efforts by the police to promote community/police relationship with the primary objective of collectively fighting crime. The police holds periodic meeting with the community. The local government should be encouraged to play a key role in such structure either as convenor or host. The local councils' halls have always served as venue for all kinds of community meetings and could serve as the venue and secretariat for the forum. The importance of taking the hosting or organisation of the forum away from the police is to encourage partnership in crime prevention rather than paternalism, where the community members are treated as mere informants. Participants in such a forum should include representatives all stakeholders in crime prevention in the community including women, non-indigenes and (Informal Police) IPS.

However, this meeting has not been consistent. It is only conveyed when there is emergency situation as observed in the cases of Niger-Delta unrest and the Boko Haram in Bornu State.

•

Impediments to The Success of Community Policing In Nigeria

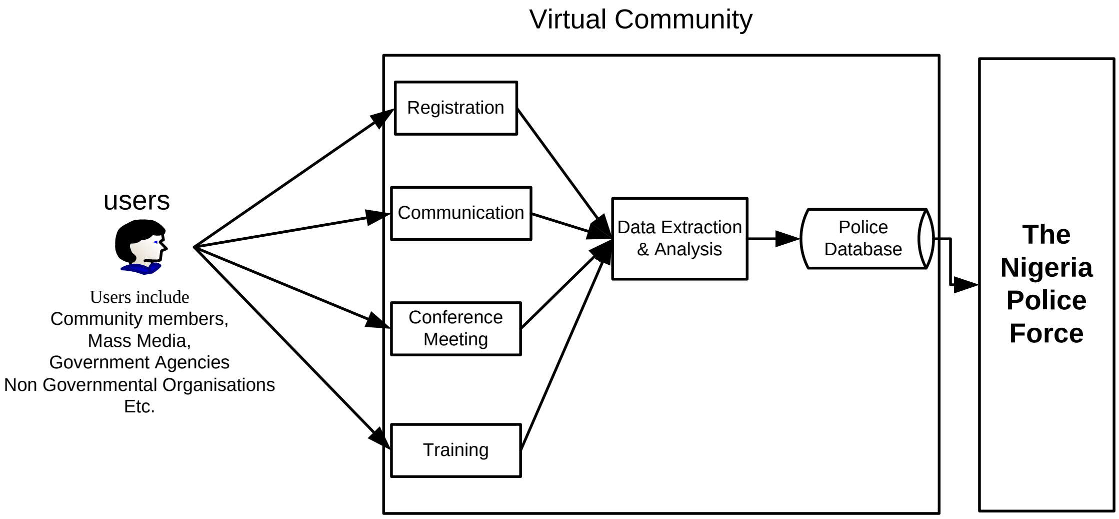
The following factors constituted impediments to the successful implementation of community policing in Nigeria.

- Internal resistance by policemen who benefited from the traditional policing and who prefer to maintain the *status quo*;
- Lack of commitment to the project by implementing officers;
- Lack of support from members of the public;
- Inadequate support from the government;
- The hostile relationship between the police and the informal policing machinery
- Poor welfare package/incentives for policemen.
- Public Attitudes towards Crime and Justice

III The Police Virtual Community Model

A community is a geographically circumscribed entity (neighborhoods, villages, etc) while A virtual community is defined as an aggregation of individuals or business partners who interact around a shared interest, where the interaction is at least partially supported and/or mediated by technology and guided by some protocols or norms[6]. Virtual communities can be dedicated to a specific topic, or they can seek to bring people with similar philosophies together. Either way, communication is digitally based, information is shared and membership is optional. Virtual communities, of course, are usually dispersed geographically, and therefore are not communities under the original definition of community. A virtual community is expected to possess the following characteristics [7]:

- It is organised around affinities or shared interests.
- It supports many to many media communication.
- The communication is graphics based supporting multimedia content(e.g. graphics, animations, video, sounds, formatted text, sound)
- No geographic boundary or physical contact.



The aforementioned features of virtual community pose greater benefits for community policing. A successful community policing requires constant interaction between the police and members of the community. It may be difficult at times to have regular physical interaction with these members at all time. It is on this note that Virtual Community model is required to bridge the communication gap between the police and members of the community.

The proposed virtual community is a web based model that will facilitate interaction between the police and members of the community. Its primary objective is to provide a platform for the police to interact with members of the public on issues of common interest such as security and safety. This interaction is intended to facilitate the fulfilment of the objective of community policing.

Three issues are central to the design of the police virtual community. These are:

i) The Virtual Work Place Environments
ii) The Services
iii) The communication tools

A. The Virtual Work Place Environment

The virtual work place environment describes the entities that constitute community members, how they are represented in the police virtual community and the access status of each. The community members' categories include:

- The Police
- Inter-security agency
- Other government agencies
- Business
- Community
- Mass Media
- CSOs/NGOs
- Foreign security agencies
- Foreign governments

The understanding is that each member can communicate with police on issue of common interest or any other issues.

However, a member must apply for membership through the virtual community before he becomes a member. The police are at the centre of the communication. All members, send their messages to the police via the virtual community. The police can initiate discussion with members. Likewise members are at liberty to start a discussion with the police.

However, provision for members to interact with one another is discouraged. The reason for this is to protect the identity of members and also protect the information supplied by members. The system is said to be centralised.

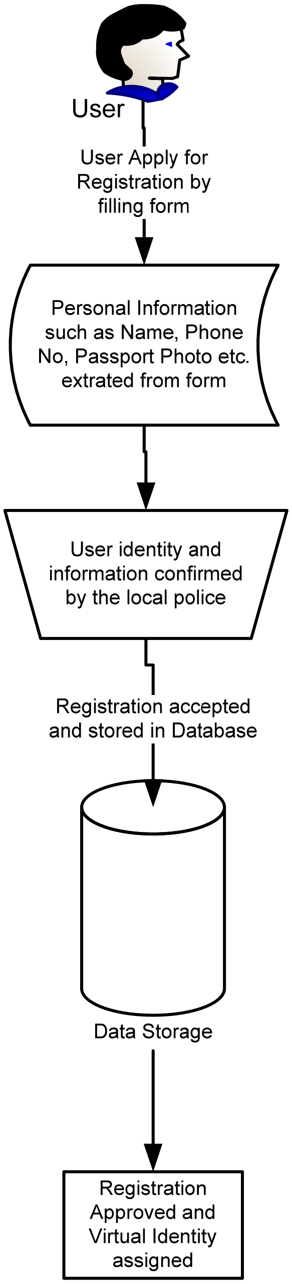
The representative of the police has administrators right. He approves members registration, coordinates discussion, store relevant information in the database and can de-member a member if situation warrants.

B. Services

It can be observed from figure one the types of interaction that can take place between the police and the community members. These are:

- **Registration**: This is required of every member before he is admitted as a member. The ideal is that if a member applies for registration, he supplies all personal details as requested by the police. The police can then use their security network to investigate the member before approval is given for membership. The diagram below describes the registration procedures.

.


User
User Apply for Registration by filling form
Personal Information such as Name, Phone No, Passport Photo etc. extrated from form
User identity and information confirmed by the local police
Registration accepted and stored in Database
Data Storage
Registration Approved and Virtual Identity assigned

Communication:- This can be inform of a discussion with the police on issues requiring urgent attention. An example is reporting a crime case, or reporting security threat in an area. The police can as well send security alert message to members of the community. Part of communication is for the police to render an account of their stewardship to the public. This will build the public's confidence in the police.

- **Meetings**: The police can organise a meeting with the virtual presence of other members of the community. The meeting agenda may problem solving issues. They can also organise a seminar on empowerment.
- **Training**: The police can organise training on security tips. They can sensitise members of the public on a newly enacted law so that the public can be aware.
- **Opinion Polls:** Polls can be conducted online on issues of common security interest. The results of such polls will aid the police decision making system.

C. The Tools

Various communications tools could be used in facilitating interaction between the police and the public. The available tools are:

- **Email**:- The Virtual environment must have provision for members to communicate through electronic mail using members virtual identity as email address. Mails can be sent even when the other member is not available online. The mail will go into its mail box and he can access it when he is available.
- **Chat**: Members can engage in real time conversation through text or voice charting

There should be a central webpage which will contain the police mission statements, important information for the public, a report of police success in various communities.

Irrespective of the communication tools available, the following are recommended for successful interaction:

- The virtual environment should support members in their decision to communicate
- It should allow users to choose among a range of communication types
- It should provide the necessary tools to initiate communication as if users are in the real world
- It should support user requirements such as use of gestures during communication mediated within the virtual world.

IV Recommendations for The Success of The Police Virtual Community.

- The Nigerian government should improve on the present infrastructural facilities such as the provision of electricity supply and communication facilities.
- Computer literacy and proficiency should be promoted among the populace.
- Computers and its accessories should be affordable
- The virtual community must be in operation round the clock i.e. 24hours in a day and seven days in a week.
- The police officer in charge of the virtual community must constantly monitor the web sites and actively participate in the citizen/police interaction
- Continous solicitation of new members will keep the site fresh and productive

IV Benefits of Virtual Community

Interaction with the police through virtual means as observed in the role of virtual communities offers a better option in crime reporting and community policing. It has some inherited benefits or advantages as highlighted below:

1. Increase community access to law enforcement information and services to the community. It can facilitate police-community dialogue, increasing transparency and enabling accurate and timely

.

information sharing that can inform police response strategies and save lives.
2. Reduce barriers to information sharing within and among law enforcement agencies across regions and across disciplines. That makes it easier to achieve multi-jurisdictional and multi-disciplinary coordinated responses to emergencies.
3. Enhance problem-solving efforts through the collection of timely and accurate data fed through robust information systems.
4. Enable standardization and access of local, state, tribal, and federal data collection and data-sharing protocols and information systems, which in turn, can enable the analysis and production of actionable intelligence.
5. Enable organizational efficiencies that inform deployment strategies, improve response times, and create opportunities for community policing activities.
6. Improve recruitment strategies and training availability, through online recruitment portals and training opportunities.

V. Conclusion

Reaping our society off crime and security hazard is a desirable factor. Consequently, justifying the need for an effective tool in Enhancing Community Policing. There might be some long term reduction in crime rates if the police were able to establish better relation with the public and increase public trust in them so that more crimes were reported. Virtual Community policing offers an effective way to have regular and constant interaction with the community members. It could also be a forum for moulding people's opinon on sensitive security and governmental issues.

The creation of a special unit of whatever designation to monitor and analysis the community interaction with the police will fit that overall aim. Such departmentalization/specialization need not be the subject of legislative but administrative action.

Additionally, the government need to provide infracstructure in various communities. At least infracstructure that will facilitate communication.

REFERENCES

[1] Centre for Law Enforcement Education in Nigeria (CLEEN) Statistics on Crime, 2003

[2] Del Buono, V. (2003) "In Search of Security", paper presented at In Search of Security: An International Conference, Canada, February 19-22, 2003.

[3] George O. S. Amadi (2011), "The Impact of Police Checkpoints on Crime and Policing in Nigeria" Faculty of Law, University of Nigeria

[4] Jemibewon, D. (2001) "The Nigerian Experience" in M. Shaw (ed.) *Crime and Policing in*

[5] *Transitional Societies*, Johannesburg: Konrad Adenauer Stiftung and South African Institute of International Affairs.

[6] Preece, Jenny (2000). "*Online Communities: Designing Usability, Supporting Sociability*". John Wiley & Sons, Chichester, UK. ISBN 0-471-80599-8

[7] Rheingold, Howard. (1994). "The Virtual Community: Homesteading on the Electronic Frontier". Addison-Wesley, Reading, MA.

[8] Scharf, W. (2000) "Community Justice and Community Policing in Post-Apartheid South Africa. How Appropriate are the Justice Systems of Africa?' Paper delivered at the *International Workshop on the Rule of Law and Development: Citizen Security, Rights and Life Choices in Low and Middle Income Countries* Institute for Development Studies, University of Sussex 1-3 June 2000.

[9] The Constitution of Federal Republic of Nigeria, 1999.

Rufai Mohammed Mutiu obtained his B.Sc degree from Ogun State University (Presently Olabisi Onabanjo University), Ago Iwoye, Ogun State, Nigeria. He got his Masters in Computer Science from University of Lagos, Akoka, Lagos, Nigeria. He is a member of Nigeria Computer Society and presently lectures at Yaba College of Technology, Lagos, Nigeria. His research area is Information Systems Design and Modelling.

Adigun Johnson Oyeranmi is a specialists in computer software, security and knowledge management. He obtained his first degree (B.Sc Computer Science) from University of Ibadan, Oyo State, Nigeria and his Masters(M.Sc. Computer Science) from University of Lagos. He is a member Nigeria Computer Society of Nigeria and Computer Professionals Council of Nigeria. He is the current Dean of The School of Technology, Yaba College of Technology, Yaba, Lagos.

.

Iterative Selective & Progressive Switching Median Filter for removal of salt and pepper noise in images

Abdullah Al Mamun
Computer Science & Engineering
Mawlana Bhashani science & Technology University
Santosh, Tangail, Bangladesh
mamun_tas_07034@yahoo.com

Md. Motiur Rahman
Computer Science & Engineering
Mawlana Bhashani science & Technology University
Santosh, Tangail, Bangladesh
mm73rahman@gmail.com

Khaleda Sultana
Computer Science & Engineering
Mawlana Bhashani science & Technology University
Santosh, Tangail, Bangladesh
khaledasultana07032@yahoo.com

***Abstract*—In this paper, we propose a new median-based switching filter, called Iterative Selective & Progressive Switching Median Filter (ISPSM), where both the noise density and threshold value are calculated dynamically from noisy input image by the noise detector, also noise detection window size is iteratively detected by noise detector. Simulation result shows that our method is significantly better than a number of existing techniques including Progressive Switching Median Filter (PSMF) in terms of image restoration and noise detection.**

Keywords-salt & pepper noise; selective & progressive switching median filter; noise detector; mean square error; peak signal to noise ratio

I. INTRODUCTION

Images are often corrupted by salt & pepper noise due to transmission errors, malfunctioning pixel elements in the camera sensors, faulty memory locations & timing errors in analog-to-digital conversion [1]. Median Filter one of the most popular filtering method has been established as a reliable method to remove noise without damaging edge details [2-4] with high computational efficiency. Several median filtering methods have been proposed for removal of salt & pepper noise densities [5-8]. The weighted median filter & center weighted median filter give more importance to current pixel preserving good image details, but offered less noise suppression when the center weighted pixel itself is corrupted [9-12]. Recently, switching schema has been studied for removal of salt & pepper noise in images [13-14]. This schema detects the noise whether the current image is corrupted by salt & pepper noise at each pixel. Then, filtering is activated for the pixels which are detected as noisy pixels, while good pixels are kept unchanged. As a switching scheme, Progressive Switching Median Filter (PSMF) [13] was proposed for removal of salt & pepper noise. In the methods of PSM filter, both the noise detector and noise filter are applied progressively. The noise detector detects a salt & pepper noise and outputs a binary flag image. The binary flag image denotes whether pixels are corrupted or not. According to a binary flag image, the filter processes to only those noisy pixels using neighborhood good pixels. Since the process of filtering according to the binary flag image, the PSM filter performs satisfactory in removing salt & pepper noise.

In this paper, we present a new median-based switching filter called Iterative Selective & Progressive Switching Median Filter (ISPSMF), where both the noise density (R), threshold value (T_D) are calculated dynamically from noisy input image by the noise detector, also noise detection window size (W_D) is iteratively detected by the noise detector, where the existing PSM filtering method manually select the value of R, T_D and W_D. In switching section of Fig. 1, if no noises are available in the image then output is the uncorrupted image and for detected noises the Iterative Noise Filter (INF) is selected. For further removal of noises (the remaining noises of the output of the INF portion), the Selective Median Filter (SMF) are applied finally. For the minimum noise filter

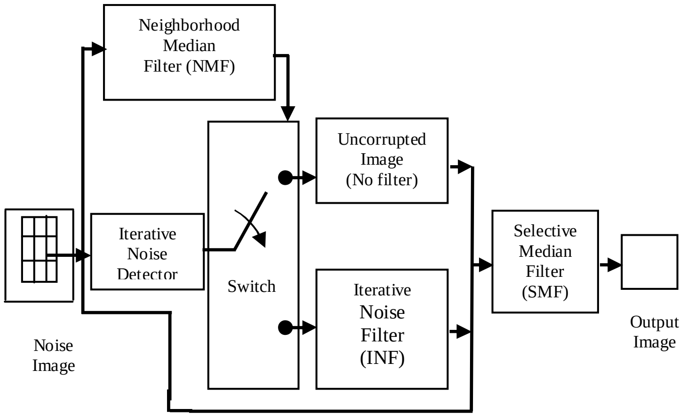

Neighborhood Median Filter (NMF)
Uncorrupted Image (No filter)
Iterative Noise Detector
Switch
Selective Median Filter (SMF)
Iterative Noise Filter (INF)
Noise Image
Output Image

Figure 1. Schematic diagram of Iterative Selective & Progressive Switching Median Filter (*ISPSMF*).

window size ($W_f \times W_f$), there may remains minimum number of noises. For that we add a Selective Median Filter (SMF), which select the noisy pixel (whose values are not 0) from the previous output of Iterative Noise Filter (INF) and replace this pixel by the neighborhood value of that's pixel.

II. THE DESIGN OF THE FILTER ISPSM

The principle of the filter ISPSM: The principle of identifying noisy pixels & processing only noisy pixels has been effective in image degradation. The limitation of the PSM filter is that defining a robust decision measure is difficult because the decision is usually based on a predefined threshold value [13]

A. Iterative Noise Detection

A noise free image should be locally smoothly varying, and is separated by edges [4]. In the n^{th} iteration (where n = 1, 2 …), for each pixel $x_i^{(n-1)}$, we first find the median value of the samples $W_D \times W_D$ (where W_D is an odd integer not smaller than 3 for better result) window centered about it. The noise density R respect to input image X is given by,

$$R=\frac{\text{sum of the pixel of X}}{(\text{Size}(X,1)*\text{size}(X,2))} \tag{1}$$

The value of noise density (R) is calculated with respect to the input image. Where,

If R≤0.25 then set the noise detection window size $W_D = 3$.

Else $W_D = 5$, Here, $W_D = 3$ is more suitable for low noise ratio & $W_D = 5$ is better for high noise ratio[5], Figure 2 with a cross point at about R = 20%.The threshold value (T_D) according to noise density(R) is given by,

$$T_D = a + (b \times R) \tag{2}$$

The Effects of T_D with respect to MSE are shown in Figure 3. According to experiment results, we choose a & b as 65, & -50, respectively. Where it appears that the best T_D is decreasing with the increase of R.[13]. Two image sequences are generated during the noise detection procedure. The first is a sequence of gray scale images, $\{\{x_{ij}^{(0)}\}, \{x_{ij}^{(1)}\}, \ldots\ldots\{x_{ij}^{(n)}\}, \ldots\ldots\}$, where the initial image $\{x_{ij}^{(0)}\}$ is the noisy image to be detected. The second is a binary flag image sequence, $\{\{f_{ij}^{(0)}\}, \{f_{ij}^{(1)}\}, \ldots\ldots\{f_{ij}^{(n)}\}, \ldots\ldots\}$, where the binary value $f_{ij}^{(n)}$ is used to indicate whether the pixel ij has been detected as a noisy pixel, i.e., $f_{ij}^{(n)} = 0$ means the pixel ij is good & $f_{ij}^{(n)} = 1$ means it has been found to be a noisy pixel. Before the first iteration, we assume that all the image pixels are good, i.e.

$$f_{ij}^{(0)} \equiv 0$$

The Progressive Switching Median Filter (PSMF) [8] where the Eq.[1-4] was introduced, but by performing some modification get best result. First median value of neighborhood pixels $m_{ij}^{(n-1)}$ is obtained using,

$$\Omega_i^W == \{(j_1,j_2) | i_1 - (W-1)/2 \le j_1 \le i_1+(W-1)/2,\; i_2 - (W-1)/2 \le j_2 \le i_2 + (W-1)/2\} \tag{3}$$

Then we get,

$$m_i^{(n-1)} = \text{Med}\{x_j^{(n-1)} \mid (j_1, j_2) \in \Omega_i^{W_D}\} \tag{4}$$

Where $\Omega_i^{W_D}$ represent the set of the pixels within a $W_D \times W_D$. Window centered about ij. And then the difference between $m_{ij}^{(n-1)}$ & $x_{ij}^{(n-1)}$ provide binary flag image $f_{ij}^{(n)}$, which is detected as a salt & pepper noise given by,

$$f_{ij}^{(n)} = \begin{cases} f_{ij}^{(n-1)}, & \text{if } |x_{ij}^{(n-1)} - m_{ij}^{(n-1)}| < T_D \\ 1, & \text{else} \end{cases} \tag{5}$$

Where T_D is calculated threshold value. Once a pixel ij is detected as a salt & pepper noise, the value of $x_{ij}^{(n)}$ is subsequently modified

$$x_{ij}^{(n)} = \begin{cases} m_{ij}^{(n-1)}, & \text{if } f_{ij}^{(n)} \neq f_{ij}^{(n-1)} \\ x_{ij}^{(n-1)}, & \text{else } f_{ij}^{(n)} = f_{ij}^{(n-1)} \end{cases} \tag{6}$$

When the noise detection procedure is stopped after the noise detection iteration number, N_D^{th} iteration, two output images { $x_{ij}^{(N_D)}$ } and { $f_{ij}^{(N_D)}$ } are obtained, but only { $f_{ij}^{(N_D)}$ } is useful for our noise filtering algorithm

B. Iterative Noise Filtering

This procedure generates a gray scale image sequence, $\{\{y_{ij}^{(0)}\}, \{y_{ij}^{(1)}\}, \ldots\ldots\{y_{ij}^{(n)}\}, \ldots\ldots\}$, is the noisy image to be filtered & a binary flag image sequence, $\{\{g_{ij}^{(0)}\}, \{g_{ij}^{(1)}\}, \ldots\ldots\{g_{ij}^{(n)}\}, \ldots\ldots\}$. Where the value $g_{ij}^{(n)} = 0$ means the pixel ij is good & $g_{ij}^{(n)} = 1$ means it is a salt & pepper noise that should be filtered. The difference between the Iterative Noise Detection & Iterative Noise Filtering (INF) procedure is that the initial flag image {$g_{ij}^{(0)}$ } of the Iterative Noise Filter (INF) procedure is not a blank image. In this method at the n^{th} iteration (n = 1, 2 …), for each pixel $y_{ij}^{(n-1)}$, Firstly we find its median value $m_{ij}^{(n-1)}$ of a $W_f \times W_f$ (W_F is an filtering window size. Also an odd integer not smaller than 3) window centered about it. Then switch the noise filter. Let M denote the number of all the pixels with $g_{ij}^{(n-1)} = 0$ in the $W_f \times W_f$ window.

If M is odd, then

$$m_{ij}^{(n-1)} = \text{Med}\{y_{ij}^{(n-1)} \mid g_{ij}^{(n-1)} = 0, ij \in \Omega_i^{W_F}\} \qquad (7)$$

Where $\Omega_i^{W_F}$ represent the set of the pixels within a $W_D \times W_D$ window centered about ij. Where,

If M is even but not 0, then

$$\begin{aligned} m_{ij}^{(n-1)} = (&\text{Med}_\text{L}\{y_{ij}^{(n-1)} \mid g_{ij}^{(n-1)} = 0, ij \in \Omega_{ij}^{W_F}\} \\ &+ \text{Med}_\text{R}\{y_{ij}^{(n-1)} \mid g_{ij}^{(n-1)} = 0, ij \in \Omega_{ij}^{W_F}\}) \div 2 \end{aligned} \qquad (8)$$

Where Med_L & Med_R denote the left ($(M/2)^{th}$ largest) & right ($(M/2+1)^{th}$ largest) median values respectively which means neighborhood pixel.

If M is greater than 0 (salt & pepper noise noisy pixel), then value $y_{ij}^{(n)}$ is modified

$$y_{ij}^{(n)} = \begin{cases} m_{ij}^{(n-1)}, & \text{if } g_{ij}^{(n)} = 0; M > 0. \\ y_{ij}^{(n-1)}, & \text{else.} \end{cases} \qquad (9)$$

Once a noisy pixel is modified, it is considered as a good pixel in the subsequent iterations

$$g_{ij}^{(n)} = \begin{cases} g_{ij}^{(n-1)}, & \text{if } y_{ij}^{(n)} = y_{ij}^{(n-1)} \\ 0, & \text{if } y_{ij}^{(n)} = m_{ij}^{(n-1)} \end{cases} \qquad (10)$$

The procedure stops after the N_F^{th} iteration when all the noisy pixels have been modified, i.e,

$$\sum_{ij} g_{ij}^{N_F} = 0 \qquad (11)$$

But there have least salt & pepper noise shown in Fig. 6, then apply selective median filter, reduce lest noise & get Restored output image {g_{ij}} of size N ×M.

III. PROPOSE (ISPSM FILTER) ALGORITHM

Step 1: Takes the pixels of the input image (Xij).

Step 2: Define the noise density(R) respect to input image (Xij) given by,

$$\text{R=}\frac{\text{sum of the pixel of X}}{(\text{Size}(\text{X},1)*\text{size}(\text{X},2))} \qquad (12)$$

Step 3: If R≤0.25 then set the salt & pepper noise detection window size (W_D) = 3 Else W_D = 5.

Step 4: Define the threshold value (T_D) according to R.

Step 5: Detect the noise respect to noise detection window size (W_D) & detection iteration number N_D (not smaller then 3 for best restoration). Also define binary flag image, f_{ij} , Where f_{ij} is define 0(zero) before first iteration.

Step 6: In iterative noise filtering (INF), respect to n^{th} iteration (n = 1, 2 …), for each pixel $y_i^{(n-1)}$. Where, $y_i^{(n-1)}$ is the gray scale image sequence. Define a binary flag image { $g_i^{(n)}$ }, the value $g_i^{(n)}$ = 0 means the pixel (i,j) is good.

Step 7: Perform Selective Median filtering (SMF) that select the noisy pixel, $g_i^{(n)}$ (whose values are not 0) and replace this value by the neighborhood value of the previous output, $y_i^{(n-1)}$ of iterative noise filter.

IV. EXPERIMENTAL RESULTS

The performance of the proposed method has been evaluated by the simulations. T_D and R are calculated dynamically from the noisy input image.

The performance of noise detection of restoration is quantitatively measured by Mean Square Error (MSE)

$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(r_{ij} - x_{ij})^2 \qquad (13)$$

The Peak Signal to Noise Ratio (PSNR)

$$PSNR = 10\log 10(255^2 / MSE) \qquad (14)$$

Where r_{ij} is the original image & x_{ij} is the restored image. The performance of the proposed method is compared with MED, CWM, PSM filters. Figure 4. & Table 1 Shows that our proposed filter reduce more noise which are applied on Bridge image. That is the PSNR of the proposed method is better than the others mentioned methods. Our proposed method is applied on lena and pepper image. Figure 5 also visually shows that its performance is better than other mentioned methods. After applying PSM method more noises are available and than our proposed method also remove them which are shown in Fig. 6, where salt & pepper noise = 0.02 for (a), (b) and salt & pepper noise = 0.31 for (c), (d).

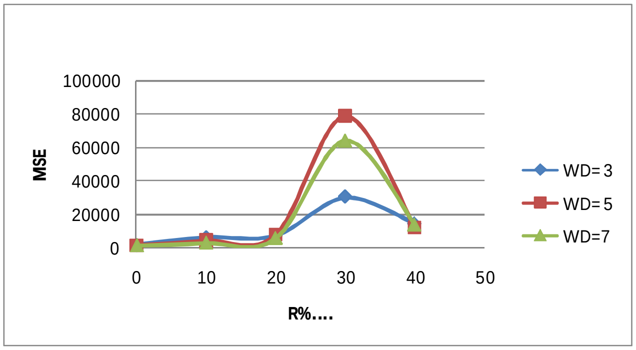


Figure 2. Effects of W_D respect to MSE.

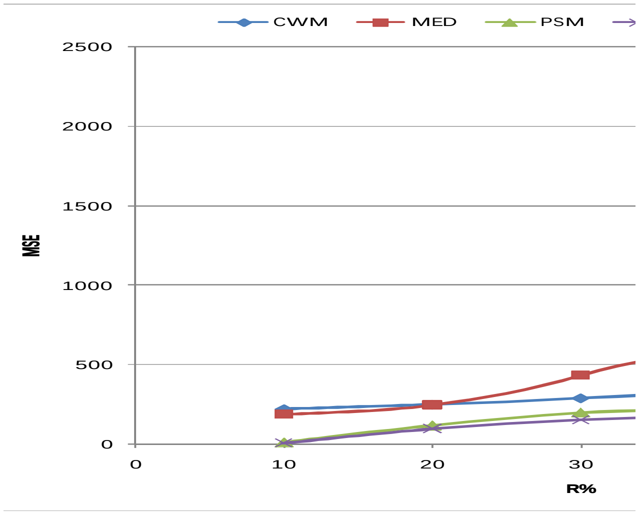


Figure 4. A comparison of different median-based filters for the restoration of corrupted image "bridge" under a large number of noise ratio.

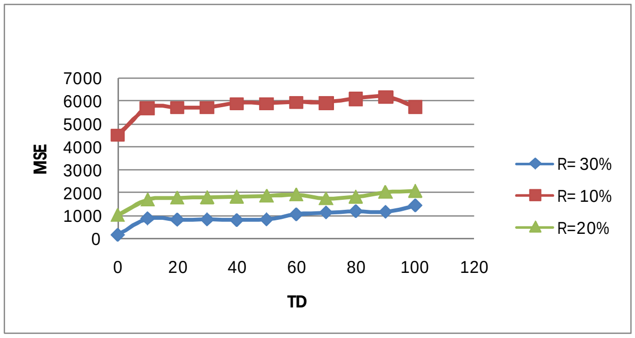
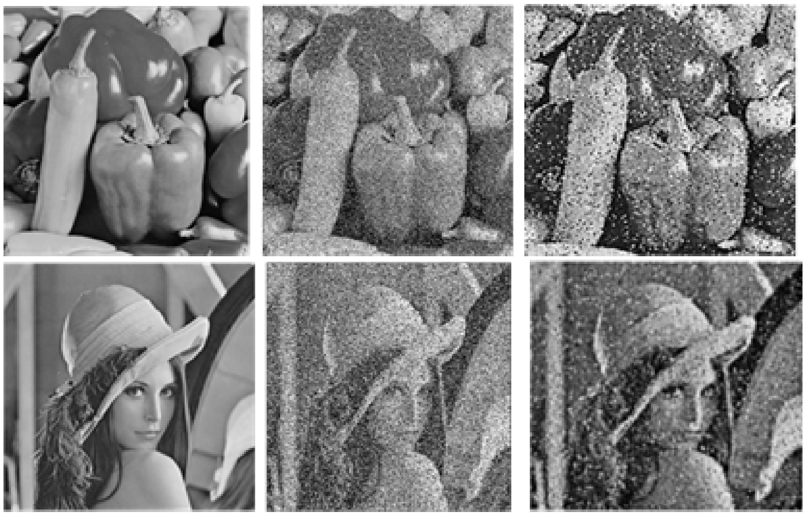


Figure 3. Effects of T_D with respect to MSE for various R

TABLE I. COMPARATIVE RESULTS OF NOISE FILTERS IN PSNR (DB) WITH SALT & PEPPER NOISE = 0.31.

	Filter Name			
	MED	CWM	PSM	ISPSM *(proposed)*
Pepper	24.134	27.881	29.384	30.631
Lena	23.310	26.991	28.092	30.039



(a) (b). (c)

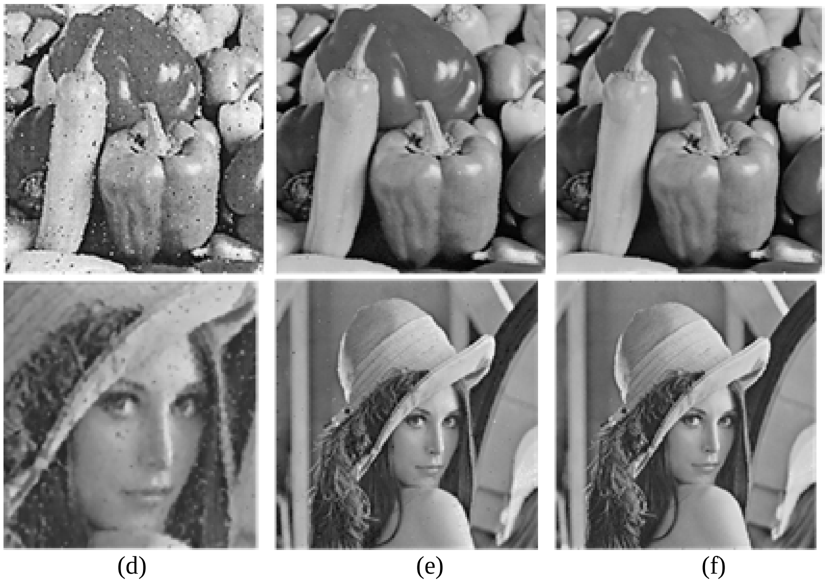
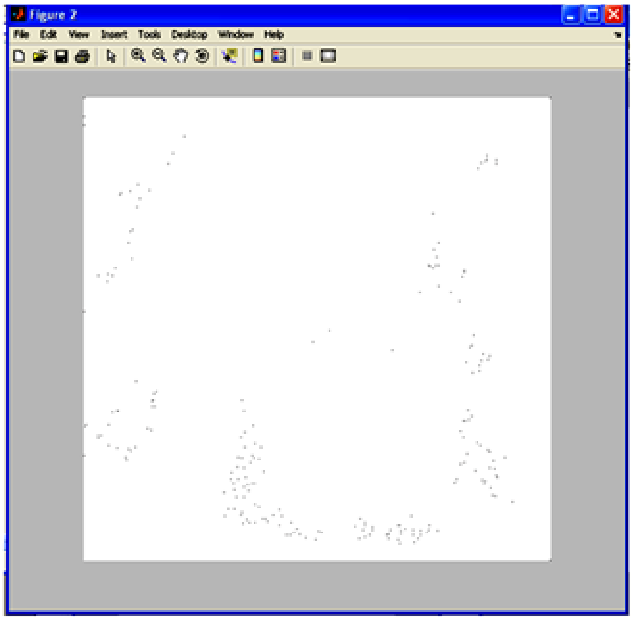
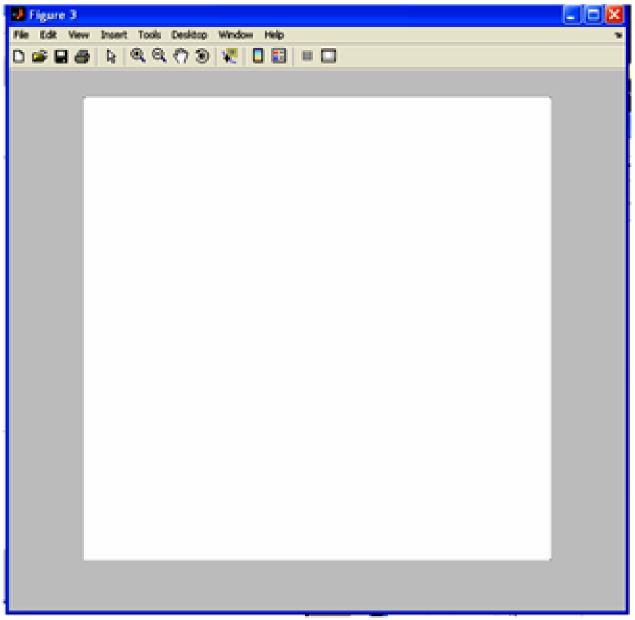


Figure 5. Visual representations of the test image with Salt & pepper noise = 0.31: (a) original image, (b) noisy image, (c) MED, (d) CWM, (e) PSM, (f) ISPSM(*proposed*).



(a)



(b)

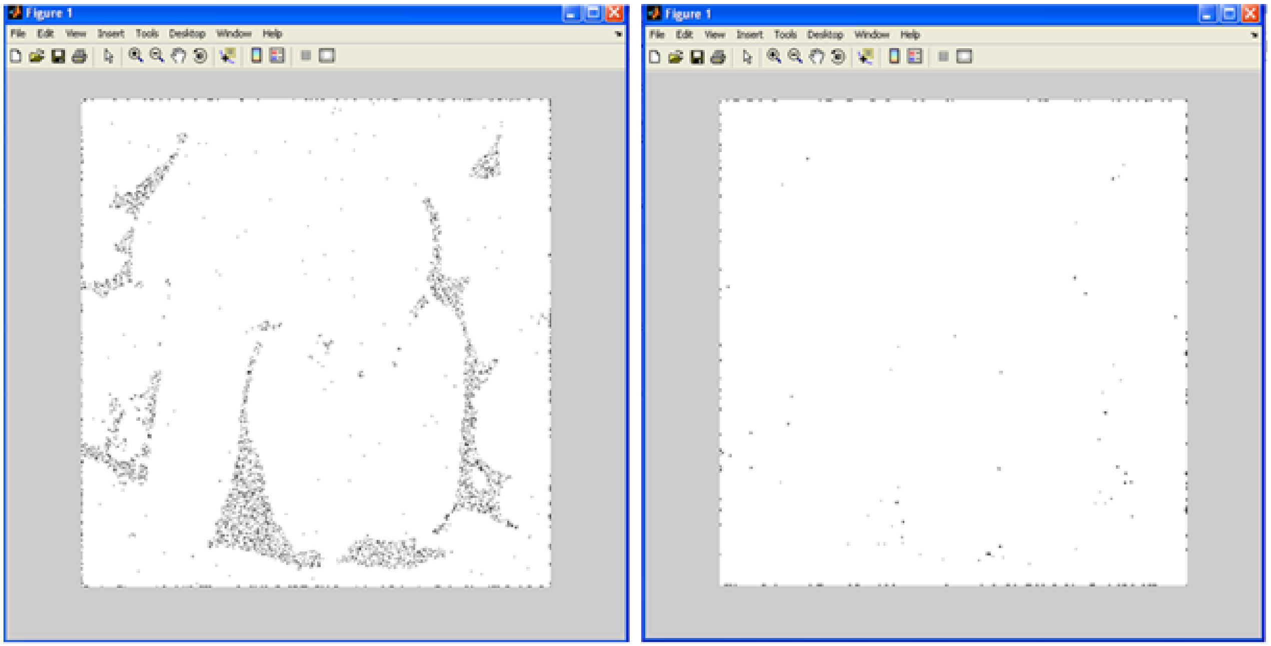


Figure 1
File Edit View Insert Tools Desktop Window Help
Figure 1
File Edit View Insert Tools Desktop Window Help

(c) (d)

Figure 6. (a) & (c) Performance of PSM filter. (b) & (d) Performance of ISPSM filter (*proposed*).

V. CONCLUSIONS

We have proposed a new median base filter that can identify more noisy pixels, also outperforms a number of existing methods (MED, CWM, PSM, ISPSM) both visually and quantitatively.

REFERENCES

[1] Gonzalez R.C. and Woods R.E., 2002. Digital Image Processing, Addison-Wesley Publishing Company.

[2] Pitas I. and Venetsanopoulos A. N., 1990. Nonlinear Digital Filters Principles and Applications, Norwell, MA: Kluwer Academic.

[3] Astola J and P.Kuosmanen,, 1997. Fundamentals of Nonlinear Digital Filtering, Boca Ratobn, CRC Pres,.

[4] W. K. Pratt,1975. Median filtering, Image Proc. Institute, University of Southern California, Los Angeles, Tech. Rep., September.

[5] N.C.Gallagher, Jr. and G.L.Wise,1981.A Theoretical analysis of the Properties of Median Filters, IEEE Trans. Acoustics, Speech and Signal Processing, vol.ASSP-29, pp.136-1141.

[6] E.Abreu, M.Lightstone, S.K.Mitra and K.Arakawa, 1996. A New Efficient Approach for the Removal of Impulse Noise from Highly Corrupted Images, IEEE Trans. Image Processing, vol.5, no.6, pp.1012-1025..

[7] T. Sun and Y. Neuvo , 1994. Detail-preserving median based filters in image processing, Pattern recognit. Lett., vol. 15, pp. 341- 347.

[8] Z. Wang and D. Zhang,, 1998. Restoration of impulse noise corrupted image using long-range correlation, IEEE Trans. Signal Processing Lett., vol.5, pp. 4-8.

[9] Xiaoyin Xu, Eric L. Miller., 2004. dongbin chen and mansoor Sarhadi Adaptive two-pass rank order filter to remove impulse noise in highly corrupted images, IEEE Transactions on Image Processing, Vol. 13, No.2.

[10] Krishnan Nallaperumal, Justin Varghese et.al, 2006.. Adaptive threshold based switching median filter for highly corrucpted images, in proc. Of CSI-IEEE First Intnl. Conf. EAIT 2006, Calcutta, India, Elsevier, pp. 103-106.

[11] Krishnan Nallaperumal, Justin Varghese et.al., 2006. Selective Switching Median Filter for the Removal of Salt & Pepper impulse noise, in proc. Of IEEE WOCN 2006, Bangalore, India.

[12] Krishnan Nallaperumal, Justin Varghese et.al., 2006 Iterative Adaptive Switching Median Filter, in proc. of IEEE ICIEA 2006, Singapore.

[13] Z. Wang and D. Zhang,, 1999. progressive switching median filter for the removal of impulse noise from highly corrupted images, IEEE Trans. Circuits Syste. II, Analog and Digit. Signal Process. , Vol. 46, No. 1, pp. 78-80.

[14] T. Chen, K.-K. Ma and L.-H. Chen., 1999. Tri-state median filter for image denoising, IEEE Trans. Image Processing, Vol. 8, pp. 1834-1838.

AUTHORS PROFILE

Abdullah Al Mamun was born in Mymensingh, Bangladesh in 1989. Currently he is the student of the department of Computer Science & Engineering in Mawlana Bhashani Science & Technology University, Santosh, Tangail, Bangladesh. His research interests include image processing & signal processing, fuzzy logic & pattern recognition, neural network, networking protocols.

Md. Motiur Rahman received the B.Eng. & M.S degree in Computer Science & Engineering from Jahangir Nagar University,Dhaka, Bangladesh, in 1995 & 2001, Where he is currently pursuing the Ph.D. degree. His research interests include digital image processing, medical image processing, computer vision & digital electronics.

Khaleda Sultana was born in Kustia, Bangladesh in 1989. Now she is the student of the department of Computer Science & Engineering in Mawlana Bhashani Science & Technology University, Santosh, Tangail, Bangladesh. His research interests include image processing.

Considering Statistical Reports of Populations Penetration in Attack to Networks

Afshin Rezakhani Roozbahani
Department of Computer Engineering
The University of Ayatollah Alozma
Boroujerdi, Boroujerd, Iran
Af.rezakhani@gmail.com

Nasser Modiri
Department of Computer Engineering
Zanjan Azad University
Zanjan, Iran
NasserModiri@yahoo.com

Nasibe Mohammadi
Department of Computer Engineering
The University of Ayatollah Alozma
Boroujerdi, Boroujerd, Iran
n.mohammadi07@gmail.com

Abstract—**because the internet traffic is increasing continuously, analyzing internet events and the penetration of countries is more important from previous years. In this article, we study the population of countries with most network traffics and consider the attacks rate that accurate in them. Also we study countries subject to attack and the rate of their attacks. These results can be used in future research to place coordinators in gorge locations of world to manage information that are passed between countries. Also these results can be used in collaborative intrusion detection systems (IDSs) for inform new attack methods to all IDSs in other location of worlds.**

Keywords-internet traffic; attacks rate; IDSs;

I. Introduction

The Internet is a global system of interconnected computer networks that use the standard Internet Protocol Suite (TCP/IP) to serve billions of users worldwide [1]. The Internet, sometimes called simply "the Net," is a worldwide system of computer networks - a network of networks in which users at any one computer can, if they have permission, get information from any other computer (and sometimes talk directly to users at other computers). It was conceived by the Advanced Research Projects Agency (ARPA) of the U.S. government in 1969 and was first known as the ARPANet. The original aim was to create a network that would allow users of a research computer at one university to be able to "talk to" research computers at other universities. A side benefit of ARPANet's design was that, because messages could be routed or rerouted in more than one direction, the network could continue to function even if parts of it were destroyed in the event of a military attack or other disaster [2]. The security disciplines of computer networks are classified into three main classes: Detection, prevention, and protection [16]. The detection methods are in charge of detecting any intrusion in networks. Prevention methods aim to deploy secure policies for underlying network(s) and finally the protection methods try to exert manager’s views for protecting the networks.

II. Internet Attack Methods

Without security measures and controls in place, our data might be subjected to an attack. Some attacks are passive, meaning information is monitored; others are active, meaning the information is altered with intent to corrupt or destroy the data or the network itself. In this section we seek the overview on the methods that are used by hackers to attack in the networks. These methods explain in below subsections [17].

A. Eavesdropping

In general, the majority of network communications occur in an unsecured or "cleartext" format, which allows an attacker who has gained access to data paths in your network to "listen in" or interpret (read) the traffic. When an attacker is eavesdropping on your communications, it is referred to as sniffing or snooping. The ability of an eavesdropper to monitor the network is generally the biggest security problem that administrators face in an enterprise. Without strong encryption services that are based on cryptography, your data can be read by others as it traverses the network.

B. Data Modification

After an attacker has read your data, the next logical step is to alter it. An attacker can modify the data in the packet without the knowledge of the sender or receiver. Even if you do not require confidentiality for all communications, you do not want any of your messages to be modified in transit. For example, if you are exchanging purchase requisitions, you do not want the items, amounts, or billing information to be modified.

C. Identity Spoofing (IP Address Spoofing)

Most networks and operating systems use the IP address of a computer to identify a valid entity. In certain cases, it is possible for an IP address to be falsely assumed— identity spoofing. An attacker might also use special programs to construct IP packets that appear to originate from valid addresses inside the corporate intranet.

After gaining access to the network with a valid IP address, the attacker can modify, reroute, or delete your data. The attacker can also conduct other types of attacks, as described in the following sections.

D. Password-Based Attacks

A common denominator of most operating system and network security plans is password-based access control. This means your access rights to a computer and network resources are determined by who you are, that is, your user name and your password.

Older applications do not always protect identity information as it is passed through the network for validation. This might allow an eavesdropper to gain access to the network by posing as a valid user.

When an attacker finds a valid user account, the attacker has the same rights as the real user. Therefore, if the user has administrator-level rights, the attacker also can create accounts for subsequent access at a later time.

After gaining access to your network with a valid account, an attacker can do any of the following:

Obtain lists of valid user and computer names and network information.

Modify server and network configurations, including access controls and routing tables.

Modify, reroute, or delete your data.

III. Considering the Population Of Contries With More Internet Traffics

A. Considering the Population of Contries

First, we study the population of some countries that play important role in internet traffics and network attacks producer. The below table is based on most network attacks producer countries. These report showing in table1 [3, 4, 5, 6, 7, 8, 9, 10].

Table 1. Population and Percentage of countries in the world

Country	*Population*	*Percentage in world*
China	1,330,141,295	19%
USA	310,232,863	4%
Netherlands	16,783,092	0.2%
Germany	82,282,988	1%
Russia	142,012,121	2%
Great Britain	62,348,447	0.9%
Canada	34019000	0.4%
Ukraine	45,415,596	0.6%
Latvia	2,231,503	0.03%
France	64,768,389	0.9%

B. Considering the Rate of Attack Producers

In this section, we study the rate of attacks that are occurred at internet. Of course our study is depended on top ten countries hosting malware [11].

Table2. Compare percentage of Contries Population with their attackers

Country	*Percentage of all attacks(hosting malware)*	*Percentage in world*
China	52.7%	19%
USA	19.02%	4%
Netherlands	5.86%	0.2%
Germany	5.07%	1%
Russia	2.58%	2%
Great Britain	2.54%	0.9%
Canada	2.22%	0.4%
Ukraine	2.17%	0.6%
Latvia	1.53%	0.03%
France	0.6%	0.9%

Of course countries with next rates are according below:

11. Spain 12. North Korea 13. Brazil 14. Cyprus 15. Sweden
16. Taiwan 17. Norway 18. Israel 19. Luxemburg 20. Estonia

Table2 compares the Percentage of all attacks (hosting malware) with Percentage of their population penetrations in world. For example, the penetration of China population in world is: 19%. Meanwhile, the hosting malware in this country is: 52.7%. This means about of 52% of world attackers, is managing their attacks in China.

C. Considering the Statistical Report of Internet Users in Above Countries

In two previous sections, we considered percentage of population and attackers. But in this section, we study the internet users at exist in these countries. This statistical report is showing as below [3].

Table 3. Considering the penetration (% population) in ten countries

Country	*Population*	*Internet*	*Penetration*

		Users	*(% Population)*
China	1,330,141,295	420,000,000	32 %
USA	310,232,863	239,232,863	77 %
Netherlands	16,783,092	14,872,200	89%
Germany	82,282,988	65,123,800	79%
Russia	142,012,121	59,700,000	43%
Great Britain	62,348,447	51,442,100	82%
Canada	34019000	26,224,900	78%
Ukraine	45,415,596	15,300,000	33%
Latvia	2,231,503	1,503,400	67%
France	64,768,389	44,625,300	69%

This table show the penetration (% population) in above countries. For example 77% of population is USA use internet in their works.

D. *Comparing above Reports*

According to internet world stats [3], total population of world is 6,845,609,960. Also according the reports of this site, total internet users in world is 1,966,514,816. Thus, the average rate of internet users in world is:

Average rate = Internet users in world / world population

Then:
Average rate = 1,966,514,816 / 6,845,609,960 = 28%

This means that from each hundred people in world, only about twenty eight of peoples work via internet to do their works. Now we consider this rate in top ten countries hosting malware. This compare is showing in table4.

Table 4. Compare population penetration factor in attacks

Country	*Percentage in world*	*Internet Users (% Population)*	*Total Internet Users in world (% Population)*	*Percentage of all attacks(hosting malware)*
China	19%	32 %	6%	52.7%
USA	4%	77 %	3%	19.02%
Netherlands	0.2%	89%	0.2%	5.86%
Germany	1%	79%	0.8%	5.07%
Russia	2%	43%	0.9%	2.58%
Great Britain	0.9%	82%	0.7%	2.54%
Canada	0.4%	78%	0.3%	2.22%
Ukraine	0.6%	33%	0.2%	2.17%
Latvia	0.03%	67%	0.02%	1.53%
France	0.9%	69%	0.6%	0.6%

This table shows the penetration of total internet users in ten countries hosting malware that are playing important role in Internet Attacks. For example, the percentage of population of China is 19% of total world population. On the other hand, 32% of the populations of this country are Internet users. Thus, about 19% * 32% = 6% of the population China is percentage of people who use Internet in all of world Internet Users. This means column4 (Total Internet Users in world (% Population)) is obtained as below:

Column4 = column2 * column3;

Figure1 show the role of penetration of populations in these countries in world attacks (hosting malware) that occurred in them.

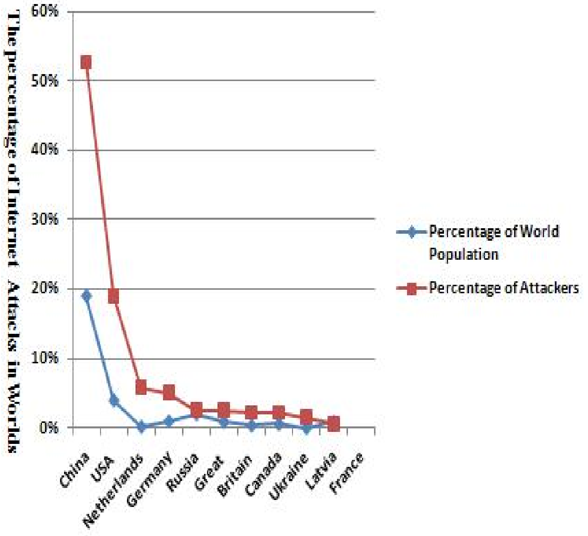
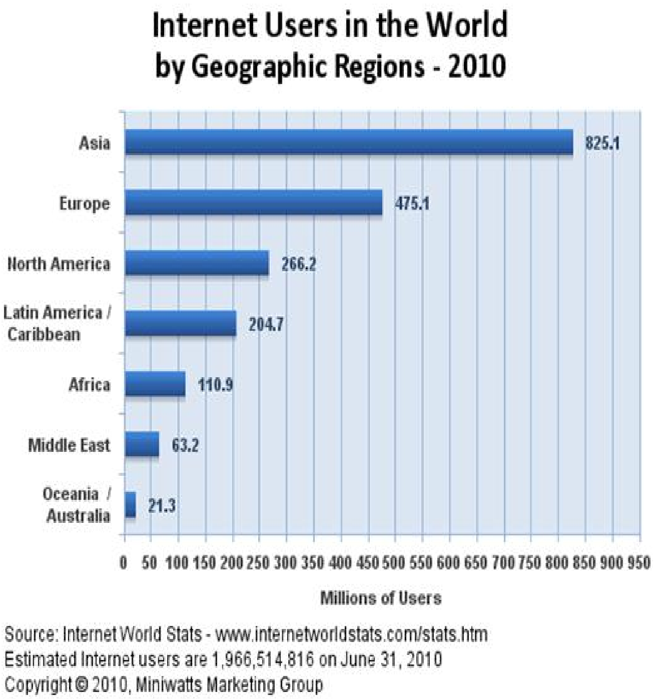


Figure 1. Relation between population and rate of malware hosting[12]

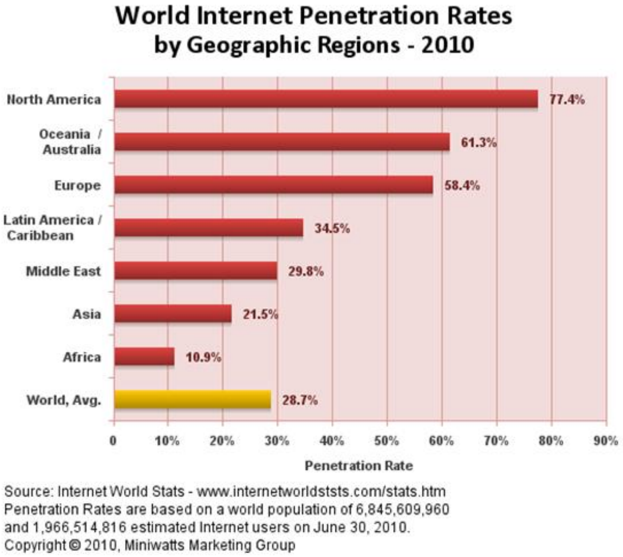
E. *Study the Internet Users in Regions*

Three below figures that are obtained by Internet World Stats [3], compare different regions by Internet Users in the world by geographic regions, world Internet penetration rates and Internet Users in the world by distribution by world regions.



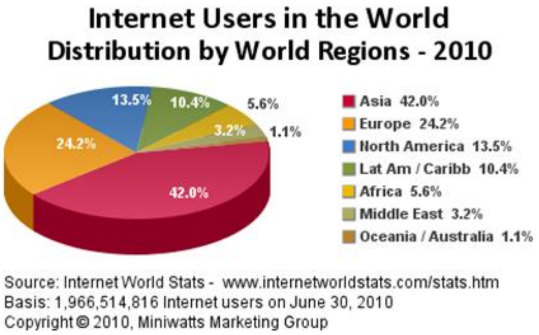
Source: Internet World Stats - www.internetworldstats.com/stats.htm
Estimated Internet users are 1,966,514,816 on June 31, 2010
Copyright © 2010, Miniwatts Marketing Group

Figure 2. Internet Users in the worlds by geographic region[12]



Source: Internet World Stats - www.internetworldststs.com/stats.htm
Penetration Rates are based on a world population of 6,845,609,960
and 1,966,514,816 estimated Internet users on June 30, 2010.
Copyright © 2010, Miniwatts Marketing Group

Figure 3. world Internet penetration rates by geographic regions[12]



Source: Internet World Stats - www.internetworldstats.com/stats.htm
Basis: 1,966,514,816 Internet users on June 30, 2010
Copyright © 2010, Miniwatts Marketing Group

Figure 4. Internet Users in the world by distribution by world regions[12]

F. Top ten malicious programs on the Internet

The twenty malicious programs most commonly used in Internet attacks are listed below. Each program has been identified more than 170,000 times and, overall, the programs listed below were involved in more than 37% (27,443,757) of all identified incidents [11].

Table 5. Top ten malicious programs on the Internet

№	*Name*	*Number of attacks*	*% of total*
1	HEUR:Trojan.Script.Iframer	9858304	13.39
2	Trojan-Downloader.JS.Gumblar.x	2940448	3.99
3	not-a-virus:AdWare.Win32.Boran.z	2875110	3.91
4	HEUR:Exploit.Script.Generic	2571443	3.49
5	HEUR:Trojan-Downloader.Script.Generic	1512262	2.05
6	HEUR:Trojan.Win32.Generic	1396496	1.9
7	Worm.VBS.Autorun.hf	1131293	1.54
8	Trojan-Downloader.HTML.IFrame.sz	935231	1.27
9	HEUR:Exploit.Script.Generic	752690	1.02
10	Trojan.JS.Redirector.l	705627	0.96

IV. Considering the reliability of Networks

Another important subject is the availability and reliability of Internet platform. For this, we study the network monitoring in some regions and ten countries hosting malware. The Internet Traffic Report monitors the flow of data around the world. It then displays a value between zero and 100. Higher values indicate faster and more reliable connections [12].

A. Internet Traffic Report in Regions

We consider in this section the score of networks in regions. The "traffic index" is a score from 0 to 100 where 0 is "slow" and 100 is "fast". It is determined by comparing the current response of a ping echo to all previous responses from the same router over the past 7 days. A score of 0 to 100 is then assigned to the current response depending on if this response is better or worse than all previous responses from that router [13]. This report shows the Global Traffic Index for the 24 hours (10/12/2010).

Table 6. Compare Internet traffics in regions

Region	*Score*	*Avg. Response Time (ms)*	*Avg. Packet Loss (%)*
Asia	68	302	9 %
Australia	83	162	0 %
Europe	75	244	11 %
North America	78	213	16 %
South America	85	144	0 %

B. Internet Traffic Report in ten Countries

We consider in this section the traffic scores in ten countries hosting malware. Similar to above subsection, this report structure is showing as below table [12].

Table 7. Compare Internet traffics in ten Countries

Country	*Score*	*Avg. Response Time (ms)*	*Avg. Packet Loss (%)*
China	96	34	0
USA	83 - 99	9 - 166	0
Netherlands	84	158	0
Germany	83	168	0
Russia	Not Consider	-	-
Great Britain	82 - 85	149 - 156	0
Canada	94	57	0
Ukraine	Not Consider	-	-
Latvia	Not Consider	-	-
France	Not Consider	-	-

V. Considering countries subject to attack

More than 86% of the 73,619,767 attacks targeted the machines of users in the ten countries listed below. This ranking has changed significantly since last year. China remains the leader in terms of numbers of potential victims, but the number of attacks dropped by 7%. Other countries which were near the top of the table last year, such as Egypt, Turkey, and Vietnam, now seem to be of less interest to cybercriminals. However, the number of attacks on users based in the US, Germany, Great Britain and Russia rose significantly [11].

Table 8. Top ten countries subject to attack in 2009

	Country	*Percentage of all attacks*
1	China	46.75%
2	USA	6.64%
3	Russia	5.83%
4	India	4.54%
5	Germany	2.53%
6	Great Britain	2.25%
7	Saudi Arabia	1.81%
8	Brazil	1.78%
9	Italy	1.74%
10	Vietnam	1.64%

VI. Our Suggested approach

A. Suggested Toplogy

We studied statistical reports from Internet traffics in some important countries and saw that the most attackers utilize these countries to networks attacks. Also they were the victim countries and subject to attack. So, if exist some powerful coordinators in these countries and strongly monitor their networks to detect/prevent attacks, other countries able work at Internet safety. This idea is showing in figure4.

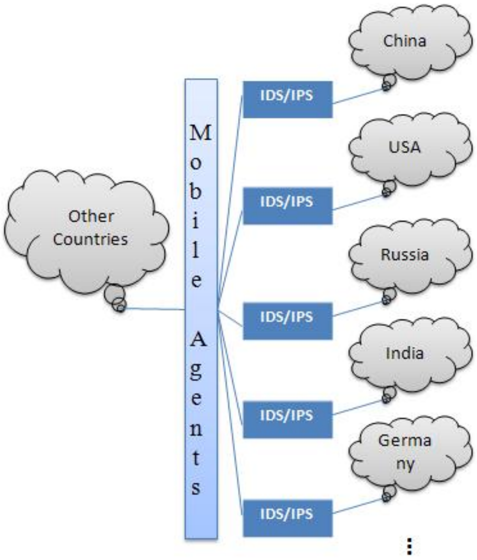


Figure 5. Placing Strong/Intelligence IDS/IPS in Countries that Subject to Attacks

Because the significant percentage of hackers, attack in few countries, we propose place powerful IDSs/IPSs to these countries. When new attack is detected by IDSs/IPSs, they send properties of detected attack to All IDSs/IPSs that exist in other countries. We evaluated this idea in other papers and showed the overhead traffic decreased by the time and do not created any significant problem [14].

Also, the relations between IDSs/IPSs can be done with secured mobile agents [15]. They propose a system where agent system will be explored on the top Grid systems that will provide security, autonomy, dynamic behavior and robust infrastructure. The key features of the proposed Agent based Grid Architecture are:

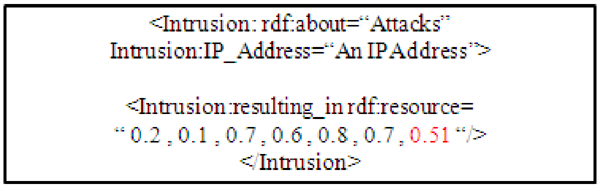
* Resuming of tasks (by using software agents) after a CPU has returned back to its idle state. All the communication and the execution of tasks are handled by software agents.

* Providing security to agents personal (confidential) data. Support of task migration is provided by our architecture due to the introduction of agents. It handles fault tolerance by maintaining multiple copies of the task.

The architecture is actually a modification of Globus Toolkit where agents are introduced. In this way we reduced the communication overhead and provided support for task migration for resource utilization [15].

B. Standardization all Detection Methods

We propose use semantic web stucture between all IDSs/IPSs to simple relation between coordinators. This work, leads to collaboration platform intrusion detection/prevention systems and causes all be abled to use from other experiences of IDSs/IPSs. We propoesd this idea is other paper Precisely. The form of semantic web that is create when an attack is detected is showing in below figure.



<Intrusion: rdf:about="Attacks"
Intrusion:IP_Address="An IP Address">

<Intrusion:resulting_in rdf:resource=
" 0.2 , 0.1 , 0.7 , 0.6 , 0.8 , 0.7 , 0.51 "/>
</Intrusion>

Figure 6. The Semantic Web Form of a detected Attack[14]

VII. CONCOLUSION

In this article, we considered the population of countries with most traffic attacks rate that accurate in them. Also we studied the probability and the rate of attacks. Studies of ten countries subject to attack in 2009 were performing. Do not found any semantic relation between population and attacks. At last, we proposed place coordinators in top countries hosting malware to detect anomalies quickly. With this, All IDSs/IPSs use from coordinators abilities to detect the attacks.

REFERENCES

[1] en.wikipedia.org/wiki/Internet.

[2] http://searchwindevelopment.techtarget.com/definition/Internet,

[3] http://www.internetworldstats.com/stats.htm

[4] http://www.indexmundi.com/netherlands/population.html

[5] http://www.countryreports.org/people/overview.aspx?Countryname=&countryId=91.

[6] http://www.trueknowledge.com/q/population_of_russia_2010

[7] www.trueknowledge.com/q/population_of_uk_2010

[8] www.statcan.gc.ca

[9] www.kyivpost.com/news/nation/detail/86668/

[10] https://www.cia.gov/library/publications/the-world-factbook/geos/fr.html.

[11] Kaspersky Security Bulletin 2009. Statistics, 2009

[12] http://www.internettrafficreport.com/

[13] http://www.internettrafficreport.com/faq.htm#trindex

[14] Afshin Rezakhani Roozbahani, L.Rikhtechi and N.mohammadi, "Converting Network Attacks to Standard Semantic Web Form in Cloud Computing Infrastructure", International Journal of Computer Applications (0975 – 8887) Volume 3 – No.4, June 2010.

[15] K.MuthuManickam, "A Security Model for Mobile Agent in Grid Environment", International Journal of Computer Applications (0975 – 8887) Volume 2 – No.2, May 2010.

[16] J. M. Kizza,"Computer Network Security", Published by Springer, 2005.

[17] Microsoft, TechNet Library, Resources for IT Professionals, http://technet.microsoft.com/en-us/library/default.aspx, Last visited at December2010

Security Implications of Ad-hoc Routing Protocols against Wormhole Attack using Random Waypoint Mobility Model in Wireless Sensor Network

Varsha Sahni
Computer Science and Engineering
Guru Nanak Dev Engineering College
Ludhiana, India.
barkhabright@gmail.com

Vivek Thapar
Computer Science and Engineering
Guru Nanak Dev Engineering College
Ludhiana, India.
vivek thapar_engg@yahoo.com

Bindiya Jain
Electronics & Communication
Engineering, DAV Institute of
Engineering & Technology, Jalandhar.
bindiyajain29@gmail.com

***Abstract*--A Wireless Sensor Network (WSN) is a network consisting of spatially distributed autonomous devices using sensors to cooperatively monitor physical or environmental conditions such as temperature, sound, vibration, pressure, motion or pollutants, at different locations.WSN is highly vulnerable to attacks because it consists of various resource-constrained devices with their low battery power, less memory, and associated low energy. Sensor nodes communicate among themselves via wireless links. However, there are still a lot of unresolved issues in wireless sensor networks of which security is one of the hottest research issues. The focus, however, has been given to the routing protocols which might differ depending on the application and network architecture. In this paper we have evaluated the affects of wormhole attack on performance of AODV and DSR routing protocols on varying node mobility. WSN's protocol has different security flaws and using these flaws many kind of attack possible on wireless sensor -network. Wormhole is one of these attacks. Wormhole attack causes serious affect on performance of the WSN protocol and preventing the attack has proven to be very difficult. In wormhole attack attacker place some malicious node in the network. A malicious node captures data packets from one location in the network and tunnels them to another malicious node at distinct location, which replays them locally. These tunnels works like shorter link in the network and so act as benefit to unsuspecting network nodes which by default seek shorter routes. This paper illustrates how wormhole attack affects performance of routing protocol in wireless sensor network using random waypoint mobility model with varying node mobility. We also analyze the effectiveness of WEP and CCMP security protocol against wormhole using DSR and AODV protocol.**

Key words:
WEP, CCMP, WSN, AODV, DSR, IMPORTANT, CBR, Random Waypoint Mobility Model

I. INTRODUCTION

Wireless Sensor Networks (WSN) is a special class of ad hoc wireless network that are used to provide a wireless communication infrastructure that allows us to instrument, observe and respond to phenomena in the natural environment and in our physical and cyber infrastructure. Sensor network [4, 6] initially consists of small or large nodes called as sensor nodes. These nodes are varying in size and totally depend on the size because different sizes of sensor nodes work efficiently in different fields.

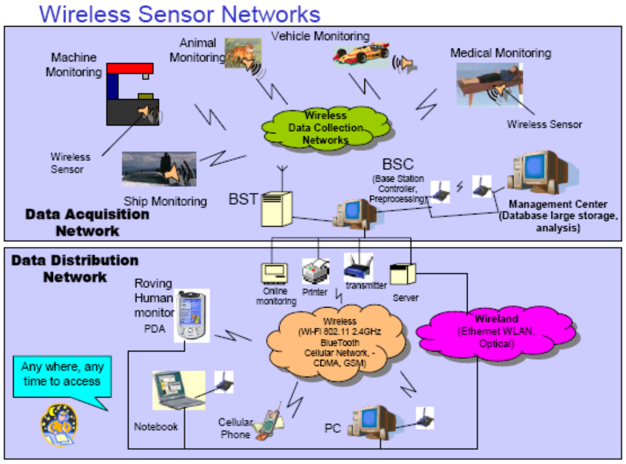

Wireless Sensor Networks
Machine Monitoring
Animal Monitoring
Vehicle Monitoring
Medical Monitoring
Wireless Data Collection Networks
Wireless Sensor
Wireless Sensor
Ship Monitoring
BST
BSC (Base Station Controller, Preprocessing)
Management Center (Database large storage, analysis)
Data Acquisition Network
Data Distribution Network
Online monitoring
Printer
transmitter
Server
Roving Human monitor
PDA
Wireless (Wi-Fi 802.11 2.4GHz BlueTooth Cellular Network, - CDMA, GSM)
Wireland (Ethernet WLAN, Optical)
Any where, any time to access
Notebook
Cellular Phone
PC

Figure-1 Wireless Sensor Network

Wireless sensor networking have such sensor nodes which are specially designed in such a typical way that they have a microcontroller which controls the monitoring, a radio transceiver for generating radio waves, different type of wireless communicating devices and also equipped with an energy source such as battery. The entire network worked simultaneously by using different dimensions of sensors [6] and worked on the phenomenon of multi routing algorithm [2] which also termed as wireless ad hoc networking.

In a wormhole attack, an attacker receives packets at one point in the network, "tunnels" them to another point in the network, and then replays them into the network from that point. For tunneled distances longer than the normal wireless transmission range of a single hop, it is simple for the attacker to make the tunneled packet arrive with better metric than a normal multihop route.

The paper is organized as follows. In the section 2, we explain the brief description of Random waypoint Mobility Model. In section 3, we explain the working of WEP and CCMP security protocols in Wireless Sensor Network. Section 4, give introduction of AODV and DSR routing protocol. Section 5, describes the security in Wireless Sensor Network. In section 6, we cover operation of wormhole attack in DSR and AODV protocols. Section 7, we discuss about the simulation setup and result of simulation and at the end in section 8, we draw the conclusion of simulation scenarios.

II. Random waypoint Mobility Model

Random waypoint model is a random-based mobility model used in mobility management schemes for mobile communication systems. This designed to describe the movement pattern of mobile user which include how their location, mobility and acceleration change over time. The Random waypoint model, first proposed by Johnson and Maltz [17], soon became a "benchmark" mobility model [20] to evaluate the Wireless Sensor Network (WSN) routing protocols, because of its simplicity and wide availability.

III. Description of Security Protocol

A. Wired Equivalent Privacy (WEP)

WEP (Wired Equivalent Privacy) was the default encryption protocol introduced in the first IEEE 802.11 standard back in 1999. It is based on the RC4 encryption algorithm, with a secret key of 40 bits or 104 bits being combined with a 24-bit Initialization Vector (IV) to encrypt the plaintext message M and its checksum – the ICV (Integrity Check Value). The encrypted message C was therefore determined using the following formula:

$$C = [\, M \,||\, \mathrm{ICV}(M) \,] + [\, \mathrm{RC4}(K \,||\, \mathrm{IV}) \,]$$

where || is a concatenation operator and + is a XOR operator. Clearly, the initialization vector is the key to WEP security, so to maintain a decent level of security and minimize disclosure the IV should be incremented for each packet so that subsequent packets are encrypted with different keys. Unfortunately for WEP security, the IV is transmitted in plain text and the 802.11 standard does not mandate IV incrimination, leaving this security measure

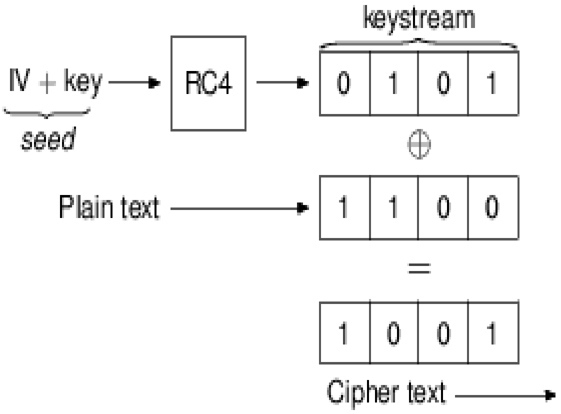


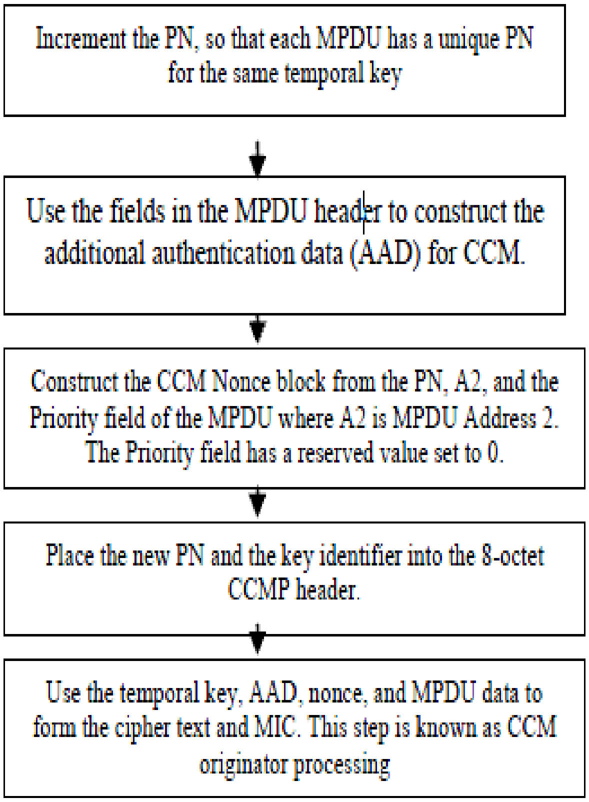
Figure 1. WEP encryption protocol

Particular wireless terminal (access point or wireless card) implementations.

B. Counter Mode with Cipher Block Chaining Message Authentication Code Protocol (CCMP)

Counter Mode with Cipher Block Chaining Message Authentication Code Protocol (CCMP) is an encryption protocol [22][23][24]. CCMP offers enhanced security compared with similar technologies such as Temporal Key Integrity Protocol (TKIP). CCMP employs 128-bit keys and a 48-bit initialization vector that minimizes the vulnerability of attack. CCMP is a robust security network association (RSNA) data confidentiality and integrity protocol. CCMP is based on the Counter Mode with CBC-MAC (CCM) of the AES encryption algorithm. CCM is a generic authenticate and encrypt block cipher mode. A unique temporal key (for each session) and a unique nonce value (a value that's used only once for each frame) are required for protecting the Medium Access Control Protocol Data Unit (MPDU). Figure3 shows CCMP encapsulation block diagram. CCMP uses a 48-bit Packet Number (PN) to protect the MPDUs. CCMP encrypts the payload of a plaintext MPDU and encapsulates the resulting cipher text using the following algorithm. Figure-2 shows CCMP encapsulation algorithm. CCMP decrypts the payload of a cipher text MPDU and decapsulates plaintext MPDU using the following algorithm. Figure 5 show CCMP decapsulation Block Diagram. Figure 4 shows CCMP decapsulation algorithm.

The decapsulation process succeeds when the calculated Message Integrity Code (MIC) matches the MIC value obtained from decrypting the received encrypted MPDU. The original MPDU header is concatenated with the plaintext data resulting from the successful CCM recipient Processing to create the plaintext MPDU.



Increment the PN, so that each MPDU has a unique PN for the same temporal key

Use the fields in the MPDU header to construct the additional authentication data (AAD) for CCM.

Construct the CCM Nonce block from the PN, A2, and the Priority field of the MPDU where A2 is MPDU Address 2. The Priority field has a reserved value set to 0.

Place the new PN and the key identifier into the 8-octet CCMP header.

Use the temporal key, AAD, nonce, and MPDU data to form the cipher text and MIC. This step is known as CCM originator processing

Figure 2 : CCMP encapsulation algorithm

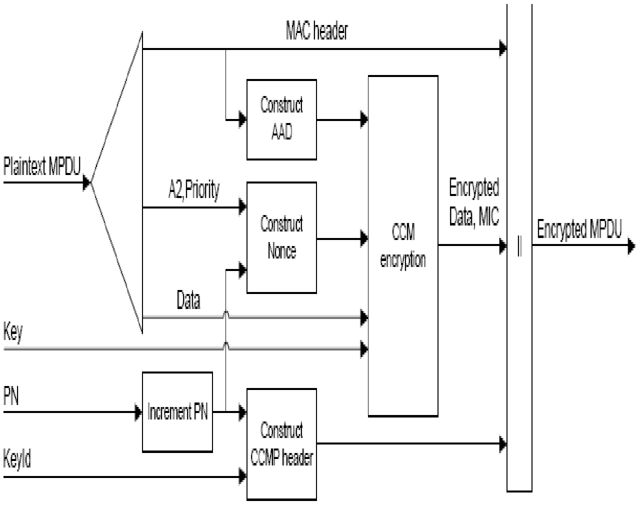
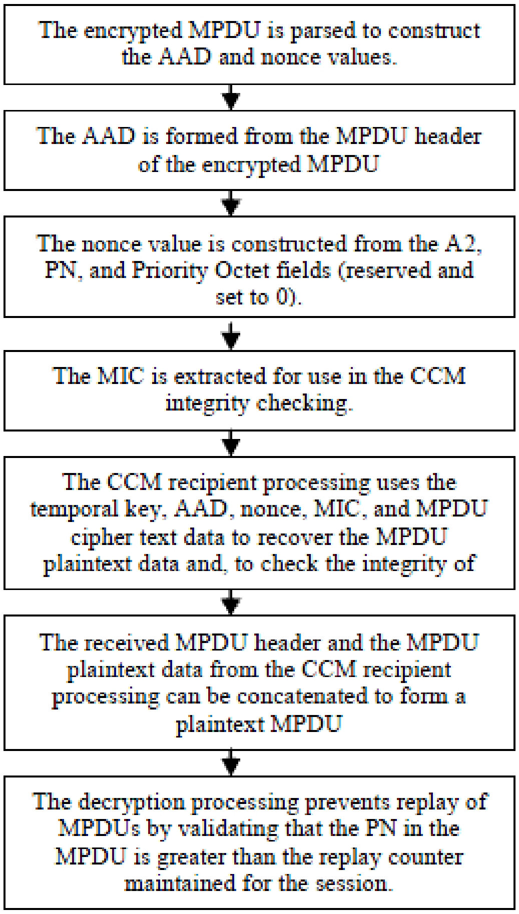


Figure 3: CCMP encapsulation Block Diagram



The encrypted MPDU is parsed to construct the AAD and nonce values.

The AAD is formed from the MPDU header of the encrypted MPDU

The nonce value is constructed from the A2, PN, and Priority Octet fields (reserved and set to 0).

The MIC is extracted for use in the CCM integrity checking.

The CCM recipient processing uses the temporal key, AAD, nonce, MIC, and MPDU cipher text data to recover the MPDU plaintext data and, to check the integrity of

The received MPDU header and the MPDU plaintext data from the CCM recipient processing can be concatenated to form a plaintext MPDU

The decryption processing prevents replay of MPDUs by validating that the PN in the MPDU is greater than the replay counter maintained for the session.

Figure 4 : CCMP decapsulation algorithm.

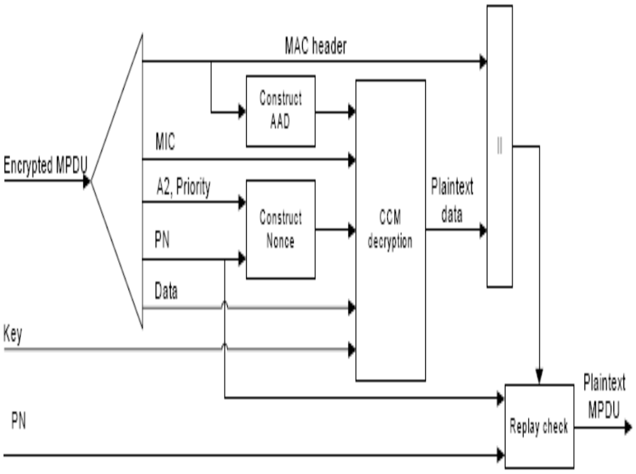


Figure 5 : CCMP decapsulation Block Diagram

IV. Description of Routing Protocol

A. Ad-Hoc on Demand Distance Vector (AODV)

AODV routing protocol [12] uses on demand approach for finding routes. In AODV the source node and the intermediate nodes store the next hop information corresponding to each flow for data packet transmission. To find a route to the destination, the source node floods the network with route request packets. The route request packets create temporary route entries for the reverse path through every node it passes in the network. When it reaches the destination a route reply is sent back through the same path the route request was transmitted. For route maintenance, every routing table entry maintains a route expiry time which indicates the time until which the route is valid. Each time that route is used to forward a data packet; its expiry time is updated to be the current time plus active route timeout. a routing table entry is invalidated if it is not used within such expiry time. AODV [7] uses an active neighbor node list for each routing entry to keep track of the neighbors that are using the entry to route data packets. These nodes are notified with route error packets when the link to the next hop node is broken. Each such neighbor node, in turn, forwards the route error to its own list of active neighbors, thus invalidating all the routes using the broken link. The main advantage of this protocol is that routes are established on demand and destination sequence numbers are used to find the latest route to the destination. the disadvantage of this protocol is that the intermediate nodes can lead to inconsistent routes if the source sequence number is very old and the intermediate nodes have a higher, but not the latest destination sequence number [3].

B. Dynamic Source Routing (DSR)

Dynamic source routing protocol (DSR) [4]: DSR is an on-demand routing protocol. The major difference between DSR and the other on demand routing protocols is that, it is beacon less and hence does not require periodic hello packets. Consider a source node that does not have a route to the destination. When it has a data packet to be sent to that destination, then it initiates a Route Request packet. This Route Request is flooded throughout the network. Each node upon receiving a Route Request broadcasts the packet to its neighbors if it has not forwarded already or if the node is not the destination node. Each Route Request carries a sequence number generated by the source node and the path it as traversed. A node, upon receiving a Route Request packet, checks the sequence number on the packet before forwarding it. The packet is forwarded only if it is not a duplicate Route Request packet. The sequence number on the packet is used to prevent loop formations and to avoid multiple transmissions of the same Route Request by an intermediate node, which receives it through multiple paths. Thus, all the nodes except the destination node, forwards a Route Request packet during the route construction phase. A destination node upon receiving the Route Request packet, replies to the source node through the reverse path the Route Request packet had traversed. Several optimization techniques have been incorporated into the basic DSR [9] protocol to improve the performance of the protocol like caching the routes at intermediate nodes. The route cache is populated with the routes that can be extracted to forward the data packet. This cache information is used by the intermediate nodes to reply to the source when they receive a Route Request packet and if they have a route to the corresponding destination.

V. Security in Wireless Sensor Network

Wireless sensor networks are complex network structures due to limitations in resources, sizes and hostile deployment environments. While implementing security many benchmarks need to be met some of these benchmarks are specific to wireless sensor networks while others are security benchmarks specific to traditional networks In following section we list various attacks[23] possible in Wireless sensor networks.

- *Denial of service attack:* A standard attack on the WSN that transmits radio signals which interfere with the radio frequencies used by the WSN, this is called “jamming”. An example of a DOS attack is when the base station is no longer able to answer the various queries.
- Sybil Attack: An attack where the adversary is able to present more than one node identity within the network. One example of such attack is when the adversary creates multiple identities of the sensor node to generate multiple readings which result in falsification of the resulted query.
- Selective Forwarding Attack: WSNs assume that each node will accurately forward the received messages. Nevertheless, if we take security into account, a compromised node may refuse to do so. It is up to the adversary that is controlling the compromised node to either forward the received readings or not. In case of not forwarding the sensor readings, the query provided by the base station may be erroneous.
- Replay Attack: In the case of a replay attack, an attacker records some traffic patterns from the network without even understanding their content and replays them later on to mislead the base station and its query answer.
- Stealthy Attack: The adversary objective in this attack is to inject false data into the network without revealing its existence. The injected false data value leads to an erroneous query result at the base station.

VI. WORMHOLE ATTACK:

Wormhole attack: In this attack, an adversary receives packets at one point in the network, tunnels them to another point in the network, and then replays them into the network from that point [20]. Malicious nodes are connected via a link called "wormhole link" using private high speed network. Wormhole attack are simple to deploy but it may cause significant damage to network.

A. Operation of wormhole attack in DSR

Wormhole attack can be carry out by using different techniques. Here we describe two methods to generate wormhole attacks in wireless sensor network. In the first type of wormhole, all packets which are received by a malicious node are duly modified, encapsulated in a higher layer protocol and dispatched to the colluding node using the services of the network nodes. These modified packets reach to colluding node just like normal node traverse form one node to another node. Once packets reach to intended malicious node, its extract the packet make the requisite modifications and send them to intended destination. In second type of attack after packets are modified and encapsulated they are send using a point to point specialized link between the malicious node. In a scenario where two malicious nodes M1and M2 are placed and they are not the immediate neighbour of source and destination nodes wormhole can be created using following steps.

Node M1 and M2 maintain a route between them using periodic update all the time. This route is use as tunnel for all other node whose traffic is routed through M1 and M2.

Whenever a ROUTE REQUEST packet is from source node S is receiving by M1 it immediately sends a route reply with minimum delay. M1 also makes the ROUTE REPLY packet (S-1-M1-M2-D) as short as possible, indicating D as an immediate neighbour of M2. Such ROUTE REPLY packets have a high probability of being selected by S as they have minimal hop-count and latency.

Node M1 inform Node M2 to initiate a route discovery process to destination node D at the mean time all packets send by S is store at M1 for a certain interval. While waiting for a route to D, if M1 receives a ROUTE REPLY packet from D to S, it verifies whether it can reach D through M2. If yes, it creates a new working source route option from M2 to D (S-M1-M2-5-D) for the buffered packets, encapsulates and sends them to M2, else it waits for the ROUTE REPLY packet to be received in response to the ROUTE REQUEST packet that was initiated by itself and M2. Upon receipt of these ROUTE REPLY packets, M1 traces an optimal route to D through M2. However, if during this waiting period, the buffer interval expires or an overflow occurs, M1 sends a ROUTE ERROR packet to S for the last received data packet.

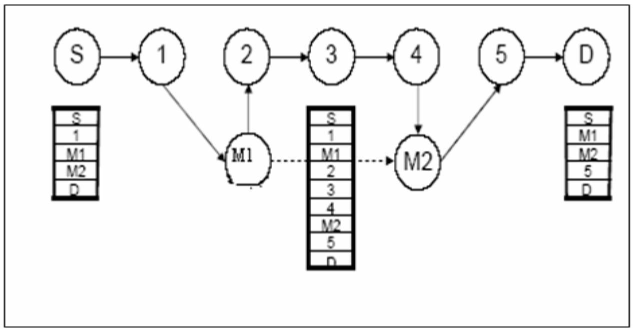


Figure-6 Wormhole attack on DSR in WSN

As an alternate mechanism, if M1 overhears any ongoing communication between S and D (S-1-2-3-4-5-D). It may initiate a new route discovery to D and also request the same through M2. Upon receipt of a route from M1 to D via M2, it can create a new Gratuitous ROUTE REPLY packet (S-1-M1-M2-D) and send it to S. Based upon the same criterion for route selections may classify the newly received route as optimal and discard the one that was already in use.

B. Operation of wormhole attack in AODV

Wormhole attack is difficult to detect. Even if the routing information is confidential, encrypted and authenticated which make is particularly very challenging in mobile ad-hoc network environment. Wormhole attack normally involve two malicious node like show in figure 7 node X and Y are the malicious node and they are attacking on traffic send by source node S to destination node D. attack start when source node S broadcast a RREP for destination node D. Since X and Y node are connected by high bandwidth wormhole link they are able to tunnel any packet between them at very high speed. so when source node S broadcast RREP it first receive by node C and D. node A broadcast this RREP packet to it neighbour node X similarly node C broadcast this packet to it neighbour node E. When malicious node X receive a RREP send by node A it tunnel the RREP packet with high speed like to node malicious node Y. Finally RREP packets receive by node D forwarded by via the path S-A-X-Y-B-D. In the same way another RREP packet is receive by node D forwarded though path S-C-E-F-G-D. However, as X and Y are connected via a high speed bus, RREQ from S-A-X-Y-B-D reaches fist to D. Therefore, destination D ignores the RREQ that reaches later and chooses D-B-A-S to unicast an RREP packet to the source node S. As a result, S chooses S-A-B-D route to send data that indeed passes through X and Y malicious nodes that are very well placed compared to other nodes in the network. Thus, a wormhole attack is not that difficult to set up, but still can be immensely harmful for a WSN

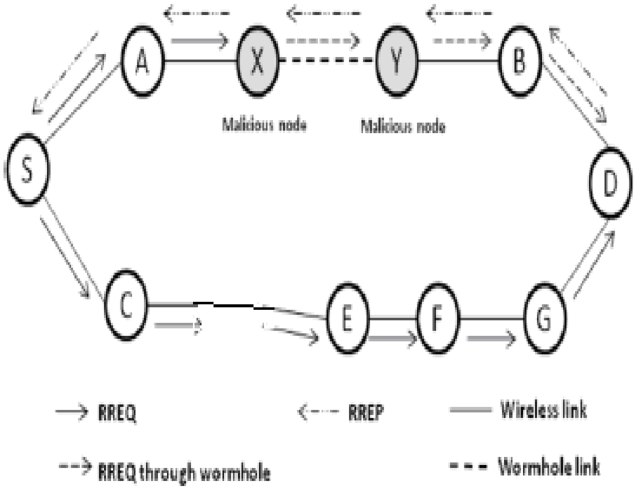


Figure-7 Wormhole attack on AODV in WSN [21]

Compare to the network that has no malicious node. But once the number of malicious node increases a particular level and it well place all over network effect of attack become severe.

VII. Simulation Setup and Result

We have used Network Simulator Qual net 5.0.2 in our evaluation. In Scenario we have place 50 nodes uniformly distributed in area of 500m x 500m. For this study, we have used random waypoint mobility model for the node movement with 0 sec pause time and 5, 10, 15, 20,25,30,35,40 meter/sec node mobility speed. The parameters used for carrying out simulation are summarized in the table 1.

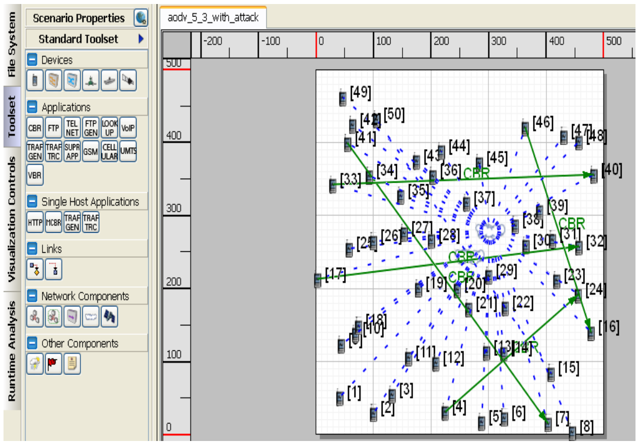


Figure 8. Simulation scenario in qualnet simulator

Parameters	*Value*
Routing Protocols	*AODV, DSR*
MAC Layer	*802.11*
Packet Size	*512 bytes*
Terrain Size	*500m * 500m*
Nodes	*50*
Mobility Model	*Random waypoint*
Data Traffic Type	*CBR*
No. of Source	*5*
Simulation Time	*200 sec.*
Node Mobility Speed	*5,10,15,20,25,30,35,40*
CBR Traffic Rate	*8 packet/sec*
Maximum buffer size for packets	*50 packets*
Security Protocols	*WEP,CCMP*

Table 1: Simulation Parameters

A. Performance Metrics:

We have used the following metrics for evaluating the performance of two on-demand reactive routing protocols (AODV & DSR):

Packet delivery ratio:

It is the ratio of data packets delivered to the destination to those generated by the sources. It is calculated by dividing the number of packet received by destination through the number packet originated from source.

$$PDF = (Pr/Ps)*100$$

Where Pr is total Packet received & Ps is the total Packet sent.

Average End-to-End Delay (second):

This includes all possible delay caused by buffering during route discovery latency, queuing at the interface queue, retransmission delay at the MAC, propagation and transfer time. It is defined as the time taken for a data packet to be transmitted across an WSN from source to destination.

$$D = (Tr – Ts)$$

Where Tr is receive Time and Ts is sent Time

Average jitter

Jitter is used as a measure of the variability over time of the packet latency across a network. A network with constant latency has no variation (or jitter). Packet jitter is expressed as an average of the deviation from the network mean latency. Jitter is cause by network congestion, timing drift, or route changes. At the sending side, packets are sent in a continuous stream with the packets spaced evenly apart. Due to network congestion, improper queuing, or configuration errors, this steady stream can become lumpy, or the delay between each packet can vary instead of remaining constant.

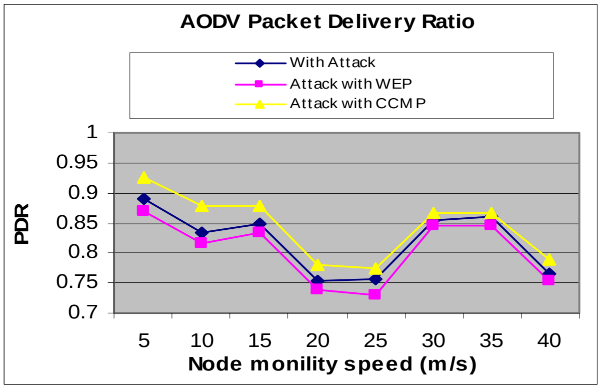


Figure 7: Packet Delivery Ratio vs. nodes mobility speed

B. AODV Packet delivery ratio under wormhole attack:

AODV protocol performance decreases as node mobility speed increase but at high node mobility (30 to 35 m/s) packet delivery ratio have improve as compare to medium node mobility speed (20 to 25 m/s). From figure 7 we can clearly see that packet delivery ratio improve when we use CCPM security protocol. WEP security has no effect on wormhole attack even packet delivery ratio decreases when we used WEP security in this scenario.

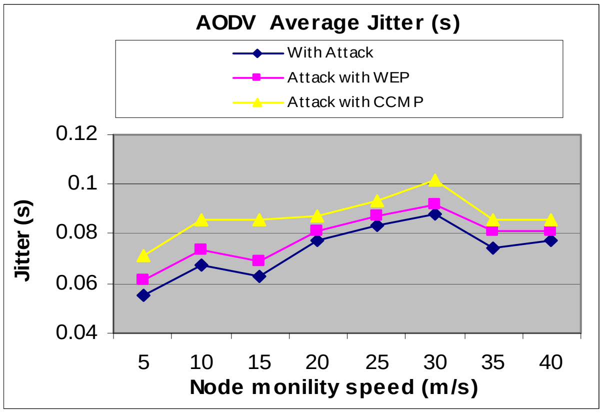


Figure 8: Average jitter vs. Nodes mobility speed

C. AODV Average Jitter under wormhole attack:

Jitter is another significant application layer parameter in mobile ad-hoc network especially in case where quality of service is required. Figure 8 show that average jitter increases when we use CCMP and WEP protocol this is because we need to perform extra step when security protocol are use like in case of CCMP we encrypt each packet using AES algorithm and WEP encrypt each packet using RC4 algorithm. These encryption algorithm take different amount of time with for different packets with add additional jitter in network.

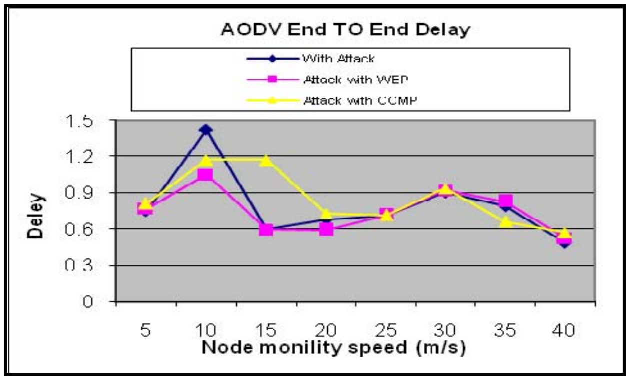


Figure 9: Average End to End-Delay vs Nodes mobility speed

D. AODV Average End to End delay under wormhole attack:

Average End to End delay increases when we use Security protocol CCMP and WEP. This is because when we introduce these protocols in existing scenario each packet needs to go through encryption and decryption process. CCMP uses AES encryption technique which is more complex then WEP used RS4 encryption so it take more time to encrypt and decrypt each packet because of this End to End delay is greater in case of CCMP as compare to WEP. From figure 9 we can observe that CCMP and WEP both Protocol are unable to improve End to End delay in case of wormhole attack.

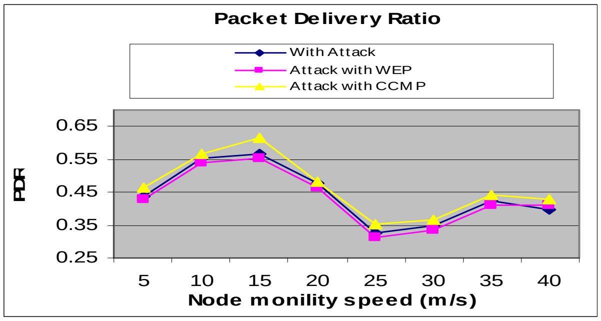


Figure 10. Packet Delivery Ratio vs. Nodes mobility speed

E. DSR Packet delivery ratio under wormhole attack:

When we compare DSR protocol performance against wormhole attack compare to ADOV we found packet delivery ratio is better for AODV. WEP protocol has almost no effect against wormhole attack for DSR protocol.

Packet delivery ratio has slightly in case of CCMP but there in no significant improvement. From figure 10 we can conclude that in case of DSR both security protocol (WEP and CCMP) are fail to prevent wormhole Attack as there is no significant improvement in packet delivery ratio.

F. DSR Average Jitter:

Average Jitter is almost double in case of DSR as compare to AODV under wormhole attack and there is no reduction in average jitter when we apply security protocol WEP and CCMP in this scenario. Figure 11 show that WEP and CCMP both security protocol are completely fail to stop wormhole attack effect of average jitter.

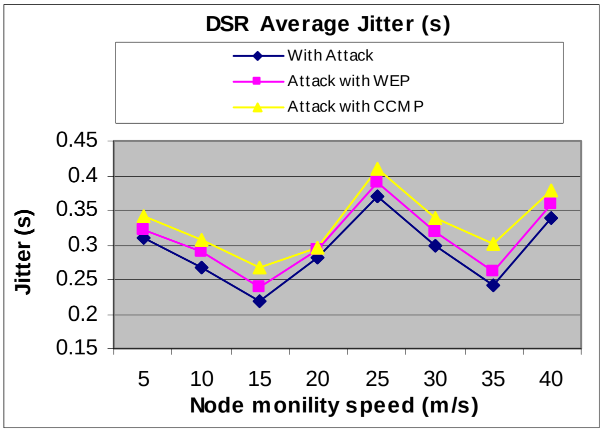


Figure 11. Average jitter vs. Nodes mobility speed.

G. DSR Average End to End delay under wormhole attack:

Wormhole attack increases the End to End delay up to 20 times as compare to no attack for DSR protocol in mobile Ad-hoc networks. Worse End to End delay is found when node mobility speed is minimum or maximum (5 or 40 m/s). Security protocol like WEP and CCMP both refailt to improve Average End to End delay.

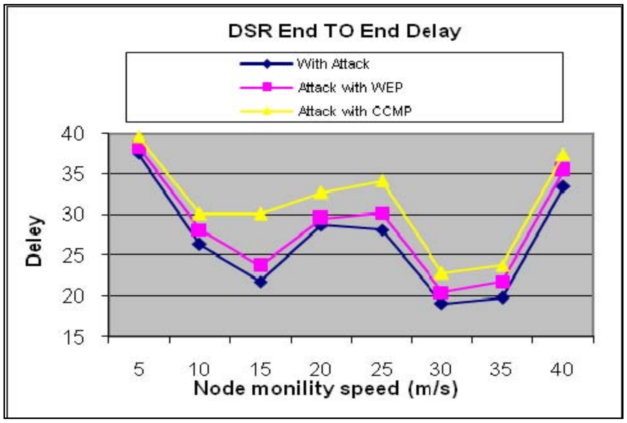


Figure 12. Average jitter vs. Nodes mobility speed.

VIII. CONCLUSION

From the figure 7 to 12, we obtain some conclusion that under wormhole attack with CBR traffic sources, AODV perform better than DSR. In case of AODV, WEP security protocol is completely fail to prevent wormhole attack but CCMP make improvement in packet delivery ratio but it fail to improve average jitter and End to End delay. DSR protocol is badly affected by wormhole. In case of DSR protocol packet delivery is ranges from 60% to 35% and End to End delay increases 20 times.

In this paper, we study the security implications that two routing protocols and two security protocols are used and their performance have been analysed against wormhole attack. This paper can be enhanced by analysing the other WSN routing protocols under different mobility model and different types of Security protocols. However, when WPA or WPA2 (802.11i) is used, the intermediate station cannot change the packet since now both the payload and the header are used for the encryption of the packet. Furthermore we propose two schemes for adjusting security (WPA or WPA2) to the new cooperative environment. In order to show feasibility of the proposed schemes, we implemented them using open source drivers.

References

[1] S. Das, C. E. Perkins, E. Royer, “Ad Hoc On Demand Distance Vector (AODV) Routing”, IETF Draft, June 2002

[2] C-K Toh “Ad Hoc Mobile Wireless Networks Protocols and Systems”, First Edition, Prentice Hall Inc, USA, 2002

[3] C.E. Perkins and E.M.Royer, “Ad-Hoc On Demand Distance Vector Routing”, Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, New Orleans, LA, USA, pages 90-100, February 1999.

[4] I. Akylidiz, W. Su, Sankarasubramaniam, and E.Cayrici, “A survey on sensor networks”, IEEE Communications Magazine, Volume: 40 Issue: 8, August 2002, pp.102-114.

[5] Fan Bai, Ahmed Helmy “A Framework to systematically analyze the Impact of Mobility on Performance of Routing Protocols for Adhoc Networks”, IEEE INFOCOM 2003.

[6] K. Akkaya and M. Younis, "A survey of Routing Protocols in Wireless Sensor Networks", Elsevier Ad Hoc Network Journal, 2005, pp 325-349.

[7] D. Johnson, Dave Maltz, Y Hu, Jorjeta Jetcheva, "The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks", Internet Draft, February 2002

[8] Suresh Kumar, R.K. Rathy and Diwakar Pandey, "Traffic Pattern Based Performance Comparison of Two Reactive Routing Protocols for Ad-hoc Networks using NS2", 2nd IEEE International Conference on Computer Science and Information Technology, 2009.

[9] D. Johnson, Y. Hu, and D. Maltz, "The Dynamic Source Routing Protocol (DSR) for Mobile", RFC 4728, Feb 2007

[10] S.Corson and J.Macker, "Routing Protocol Performance Issues and Evaluation considerations", RFC2501, IETF Network Working Group, January 1999.

[11] S. R. Biradar, Hiren H D Sharma, Kalpana Shrama and Subir Kumar Sarkar, "Performance Comparison of Reactive Routing Protocols of WSNs using Group Mobility Model", IEEE International Conference on Signal Processing Systems, pages 192-195 2009.

[12] C. Perkins, E. Belding-Royer, S. Das, quet, "Ad hoc On-Demand Distance Vector (AODV) Routing", RFC 3561, July 2003

[13] N.Aschenbruck, E.Gerhands-Padilla, P.Martini, "A Survey on mobility models for Performance analysis in Tactical Mobile networks," Journal of Telecommunication and Information Technology,Vol.2 pp.54-61,2008

[14] X. Hong, M. Gerla, G. Pei, and C.-C. Chiang, "A group mobility model for ad hoc wireless networks," in *ACM/IEEE MSWiM*, August 1999.

[15] http://www-scf.usc.edu/~fbai/important/, referred on February 2010.

[16] http://nile.usc.edu/important/, referred on February 2010.

[17] Bai, Fan; Helmy, Ahmed (2006). A Survey of Mobility Models in Wireless Adhoc Networks. (Chapter 1 in Wireless Ad-Hoc Networks. Kluwer Academic. 2006.

[18] Broch, J; Maltz DA, Johnson DB, Hu Y-C, and Jetcheva J (1998). "A performance comparison of multi-hop wireless ad hoc network routing protocols". roceedings of the Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking(Mobicom98), ACM, October 1998.

[19] A. A. Pirzada and C. McDonald, "Kerberos assisted authentication in mobile ad-hoc networks, in Proceedings of the 27th Australasian Computer Science Conference (ACSC), 2004

[20] A. Perrig, Y. C. Hu, and D. B. Johnson, Wormhole Protection in Wireless Ad Hoc Networks, Technical Report TR01-384, Department of Computer Science, Rice University, 2001.

[21] Rashid Hafeez Khokhar, Md Asri Ngadi and Satira Mandala, "A Review of Current Routing Attacks in Mobile Ad Hoc Networks", International Journal of Computer Science and Security, pp. 18-29, Volume-2 Issue-3

[22] Changhua He and John C Mitchell, "Security Analysis and Improvements for IEEE 802.11i", in the Proceedings of the 12th Annual Network and Distributed System Security Symposium (NDSS'05), 2005.

[23] H. Lan Nguyen and U, Trang Nguyen "A study of different types of attacks on multicast in mobile ad hoc networks", Ad Hoc Network, VoI.6, No. I,2007

[24] Specification for the Advanced Encryption Standard (AES), FIPS 197, U.S. National Institute of Standards and Technology.

[25] I. Akylidiz, W. Su, Sankarasubramaniam, and E.Cayrici, "A survey on sensor networks", IEEE Communications Magazine, Volume: 40 Issue: 8, August 2002, pp.102-114.

Vivek Thapar did his graduation from Punjab Technical University, Kapurthala and post graduation from Punjabi University Patiala with 72%. He involved in research since last four years. His research paper has been published in different national and international journals. He presented many papers in different seminar and conferences. Currently He involved developing novel software for different statistical methods and presently working as a assistant professor in Computer Science and Engineering department at Guru Nanak Dev Engineering College, Ludhiana, India. His Area of Specializations is Network Security and Web Technologies. He is currently doing PhD from Punjab Technical University.

Bindiya Jain did her graduation from GNDU, Amritsar and her M. Tech. from Electronics and Communication Engineering at DAV Institute of Engineering & Technology, Jalandhar. Her research paper has been published in different national and international journals. She presented many papers in different seminar and conferences. Her area of interest is wireless sensor network. She is currently doing phd in wireless sensor network from Punjab Technical University.

Varsha Sahni has received B-Tech degree from Punjab Technical University, Jalandhar in 2009 and pursuing her M-Tech Degree in Computer Science and Engineering at Guru Nanak Dev Engineering College, Ludhiana, India from Punjab Technical University, Jalandhar. Her research interests are in the fields of Routing Algorithms, Routing Protocols Load Balancing and Network Security, wireless sensor network. She has published many national and international and international journal papers.

An Empirical Comparison of Boosting and Bagging Algorithms

R. Kalaichelvi Chandrahasan
College of Computer Studies
AMA International University
Kingdom of Bahrain
kalai_hasan@yahoo.com

Angeline Christobel Y
College of Computer Studies
AMA International University
Kingdom of Bahrain
angeline_christobel@yahoo.com

Usha Rani Sridhar
College of Computer Studies
AMA International University
Kingdom of Bahrain
ama_usharani@yahoo.com

Arockiam L
Dept.of Computer Science
St. Joseph's College
Tiruchirappalli, TN, India
larockiam@yahoo.co.in

Abstract - **Classification is one of the data mining techniques that analyses a given data set and induces a model for each class based on their features present in the data. Bagging and boosting are heuristic approaches to develop classification models. These techniques generate a diverse ensemble of classifiers by manipulating the training data given to a base learning algorithm. They are very successful in improving the accuracy of some algorithms in artificial and real world datasets. We review the algorithms such as AdaBoost, Bagging, ADTree, and Random Forest in conjunction with the Meta classifier and the Decision Tree classifier. Also we describe a large empirical study by comparing several variants. The algorithms are analyzed on Accuracy, Precision, Error Rate and Execution Time**.

Key Wrods - Data Minig, Classification, Meta classifier, Decision Tree

I. INTRODUCTION

Data Mining is an iterative and multi step process of knowledge discovery in databases with the intention of uncovering hidden patterns. The huge amount of data to process is more and more significant in the world. Modern data-mining problems involve streams of data that grow continuously over time that includes customer click streams, telephone records, large sets of web pages, multimedia data, sets of retail chain transactions, assessing credit risks, medical diagnosis, scientific data analysis, music information retrieval and market research reports [32].

Classification algorithm is a robust data mining tool that uses exhaustive methods to generate models from a simple to highly complex data. The induced model is used to classify unseen data instances. It can be referred as supervised learning algorithms because it assigns class labels to data objects. There are many approaches to develop the classification model including decision trees, meta algorithms, neural networks, nearest neighbor methods and rough set-based methods [14, 17].

The Meta classifiers and the decision trees are the most commonly used classification algorithms, because of their ease of implementation and easier to understand compared to other classification algorithms.

The main objective of this paper is to compare AdaBoost, Bagging, ADTree and Random Forest algorithms which use bagging or boosting techniques based on Accuracy, Precision, Error Rate and Processing Time. The implementations of these algorithms were taken place on three different medical datasets, "Wisconsin-BreastCancer", "Heart-statlog" and "Liver-disorders" obtained from UCI Machine Learnig Repository [40].

Section 2 presents the proposed ensemble methods for building ensembles that are based on bagging and boosting techniques, while section 3 discusses the procedure for performance estimation. Experiment results using three medical data sets and comparisons of performance attributes such as accuracy, precision, error rate and the processing time with four algorithms are presented in section 4. We conclude in section 5 with summary and further research areas.

II. BOOSTING AND BAGGING APPROACHES

Meta Learning is used in the area of predictive data mining, to combine the predictions from multiple models. It is significantly useful when the types of models are very different in their nature. In this perspective, this method is defined as Stacking or Stacked Generalization. The predictions from various classifiers can be used as input to a meta-learner. The final best predicted classification will be created in combining the predictions from the multiple methods. This procedure yields more accurate predictions than any other classifiers.

Decision tree induction is a data mining induction techniques to solve the classification problems. The goal in constructing a decision tree is to build a tree with accuracy and better performance. It is made of root, nodes, branches, and leaf nodes. The tree is used in classifying unknown data records. To classify an instance, one starts at the root and finds the branch corresponding to the value of that attribute observed in the instance. This process is repeated at the sub tree rooted at that branch until a leaf node is reached. The resulting classification is the class label on the leaf [26].

In this paper we study the classification task with more emphasis on boosting and bagging methods classification. The four popular ensemble algorithms are boosting, bagging, rotation forest and random subspace method. This paper describes the boosting and bagging techniques. Boosting induces the ensemble of weak classifiers together to create one strong classifier. In boosting successive models give extra weights to the earlier predictors. While In bagging, successive trees do not depend on earlier trees. Each model is independently constructed using a bootstrap sample of the data set. In the end, overall prediction is made by majority voting. The paper concludes with two novel classifiers Meta classifier and Decision Trees classifier that give idea of their Accuracy and Precision attributes.

A. *Meta Classifier: AdaBoost Algorithm*

Adaptive boosting is a popular and powerful meta ensemble algorithm. "Boosting" is an effective method for the improvement in the performance of any learning algorithm. It is also referred as "stagewise additive modeling". The model is a more user friendly algorithm. The algorithm does not suffer from overfitting. It solves both the binary classification problems as well as multiclass problems in the machine learning community. AdaBoost also gives an extension to regression problems. Boosting algorithms are stronger than bagging on noise free data. The algorithm depends more on data set than type of classifier algorithms. The algorithm puts many weak classifiers together to create one strong classifier. It is a sequential production of classifiers.

To construct a classifier:

1. A training set is taken as input
2. A set of weak or base learning algorithms are called repeatedly in a series of rounds to maintain a set of weights over the training set. Initially, all weights are set equally, but on each round, the weights of incorrectly classified examples are increased so that the weak learner is forced to focus on the hard examples in the training data.
3. This boosting can be applied by two frameworks, i) boosting by weighting ii) boosting by sampling. In boosting by weighting method, the base learning algorithms can accept a weighted training set directly. With such algorithms, the entire training set is given to the base learning algorithm. And in boosting by sampling examples are drawn with replacement from the training set with probability proportional to their weights.
4. The stopping iteration is determined by cross validation.

The algorithm does not require prior knowledge about the weak learner and so can be flexibly combined with *any* method for finding weak hypotheses. Finally, it comes with a set of theoretical guarantees given sufficient data and a weak learner that can reliably provide only moderately accurate weak hypotheses.

The algorithm is used on learning problems having either of the following two properties. The first property is that the observed examples tend to have varying degrees of hardness. The boosting algorithm tends to generate distributions that concentrate on the harder examples, thus challenging the weak learning algorithm to perform well on these harder parts of the sample space. The second property is that the algorithm is sensitive to changes in the training examples so that significantly different hypotheses are generated for different training sets.

B. *Meta Classifier: Bagging Algorithm*

Bagging is a machine learning method of combining multiple predictors. It is a model averaging approach. Bagging is a technique generating multiple training sets by sampling with replacement from the available training data. It is also known as bootstrap aggregating. Bootstrap aggregating improves classification and regression models in terms of stability and accuracy. It also reduces variance and helps to avoid overfitting. It can be applied to any type of classifiers. Bagging is a popular method in estimating bias, standard errors and constructing confidence intervals for parameters.

To build a model,

i) split the data set into training set and test set.
ii) Get a bootstrap sample from the training data and train a predictor using the sample.

Repeat the steps at random number of times. The models from the samples are combined by averaging the output for regression or voting for classification. Bagging automatically yields an estimate of the out of sample error, also referred to as the generalization error. Bagging works well for unstable learning algorithms like neural networks, decision trees and regression trees. But it works poor in stable classifiers like k-nearest neighbors. The lack of interpretation is the main disadvantage of bagging. The bagging method is used in the unsupervised context of cluster analysis.

C. *Decision Tree Classifier: ADTree Algorithm*

The Alternating Decision Tree (ADTree) is a successful machine learning classification technique that combines many decision trees. It uses a meta-algorithm boosting to gain accuracy. The induction algorithm is used to solve binary classification problems. The alternating decision trees provide a mechanism to generate a strong classifier out of a set of weak classifier. At each boosting iteration, a splitter node and two prediction nodes are added to the tree, to generate a decision tree. In accordance with the improvement of purity, the algorithm determines a place for the splitter node by analyzing all prediction nodes. Then the algorithm takes the sum of all prediction nodes to gain overall prediction values. A positive sum represents one class and a negative sum represents the other in two class data sets. A special feature of ADTree is the trees can be merged together. In multiclass problems the alternating decision tree can make use of all the weak hypotheses in boosting to arrive at a single interpretable tree from large numbers of trees.

D. *Decision Tree Classifier: Random Forest Algorithm*

A random forest is a refinement of bagged trees to construct a collection of decision trees with controlled variations. The method combines Breiman's bagging and Ho's random subspace method. The algorithm improves on bagging by de-correlating the trees. It grows trees in parallel independently of one another. They are often used in very large datasets and a very large number of input variables. A random forest model is made up of hundreds of decision trees. It does not require tree pruning and it handles continuous and categorical variables and missing values. The algorithm can be used to generate tree-base clusters through sample proximity.

The Random Forest algorithm is as follows:

1. First Randomization (Bagging)
Random Forest uses Bootstrap aggregation / bagging method of ensemble learning that uses bootstrap sample (i.e sampling with replacement from the original data) with a randomized selection of features at each split in tree induction. Grow an un-pruned tree with this bootstrap. Splits are chosen by purity measures, Classification uses Gini or deviance, while regression uses squared error.
2. Second Randomization (Selection of subset Predictors)
At each internal node, randomly select the best among a subset of predictors and determine the best split.
$m_{try\,-}$ number of predictors to try at each split.
k – total number of predictor
For classification $m_{try} = \sqrt{K}$
for Regression =k/3

Bagging is a special case of Random Forest where $m_{try=}$ k

Subset of predictors is much faster to search than all predictors. The overall Prediction is made by majority voting (classification) or averaging (regression) the predictions of the ensemble. As it is parallel algorithm type, several random forests can be run on many machines and then aggregate the votes component to get the final result. As it has only two parameters i) the number of variables in the random subset ii) and the number of trees in the forest, it is user-friendly.

For each tree grown, 33-36% samples are not selected in the bootstrap, called "Out Of Bootstrap" or "Out of Bag" (OOB) samples [8]. Predictions are made using these OOB samples as input. OOB estimate of error rate will be computed by aggregating the OOB predictions. As it generates an internal unbiased estimate of the test error, cross validation is not necessary. The algorithm builds trees until the errors no longer decreases. The number of predictors determines the number of trees necessary for good performance.

III. PERFORMANCE EVALUATION

Performance evaluation is a significantly important factor of any classifier. Performance evaluation includes the performance metrics for evaluating a single classifier, the metrics for comparing multiple classifiers and measure for the effectiveness of the classifiers, which is the ability to take the right classification decisions. Various performance metrics are used for classification effectiveness evaluation, including accuracy, correct rate, recognition rate, error rate, false rate, reject rate, recall and precision.

Cross validation is considered as a standard procedure for performance estimation. There are several approaches in cross validation methods such as Resubstitution Validation, Hold-out Validation, k-fold cross validation, Leave-One-Out cross-validation and Repeated k-fold cross-validation. In this study, we have selected k-fold cross validation for evaluating the classifiers [3, 9].

The estimations of accuracy, precision and error rate are the key factors to determine the algorithms' effectiveness in a supervised learning environment. In our empirical tests, these characteristics are evaluated using the data from the confusion matrix obtained. A confusion matrix contains information about actual and predicted classifications obtained by a classification algorithm. The time taken to build the model is also taken as another factor for the comparison.

The Accuracy, Precision and the Error are computed as follows:

Accuracy = (a+d)/(a+b+c+d)
Precision = (d)/(b+d)
Error = (b+c)/(a+b+c+d)

Where,

- *a* is the number of correct predictions that an instance is negative,
- *b* is the number of incorrect predictions that an instance is positive,
- *c* is the number of incorrect of predictions that an instance negative, and
- *d* is the number of correct predictions that an instance is positive.

IV. EXPERIMENTAL ANALYSIS

We carried out some experiments using Wisconsin-Breast Cancer, Heart-statlog and Liver-disorders data sets attained from the UCI Machine Learning Repository [40]. In our comparison study, the implementations of algorithms were done by a machine learning algorithm tool Weka version 3.6.5. Weka is a very supportive tool in learning the basic concepts of data mining where we can apply different options and analyze the output that is being produced.

Table 1 shows the datasets used for the implementation of algorithms with their number of instances, the number of attributes.

Table 1: Description of the Datasets

Dataset	Instances	Attributes
Wisconsin-BreastCancer	699	10
Heart-statlog	270	14
Liver-disorders	345	7

Table 2 shows the accuracy of various classifiers. The Figure 1 gives an idea about the accuracy of the selected algorithms in graphical format.

Table 2: Accuracy Comparison

Dataset	Accuracy (%)			
	Meta Classifier		Decision Tree	
	AdaBoost	Bagging	ADTree	Random Forest
Wisconsin-BreastCancer	94.85	95.57	95.85	96.14
Heart-statlog	80.0	78.89	78.52	78.15
Liver-disorders	66.09	71.3	59.71	68.99

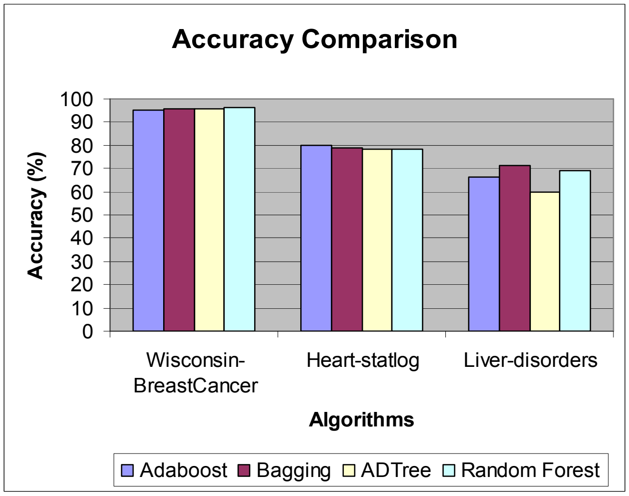


Figure 1: Graphical Representation of Accuracy

The precision comparison among the four algorithms is shown in Table 3 and the graphical representation can be seen in Figure 2.

Table 3: Precision Comparison

Dataset	Precision (%)			
	Meta Classifier		Decision Tree	
	AdaBoost	Bagging	ADTree	Random Forest
Wisconsin-BreastCancer	92.89	92.34	94.17	93.5
Heart-statlog	77.5	77.39	75.83	76.52
Liver-disorders	67.36	72.25	65.02	73.85

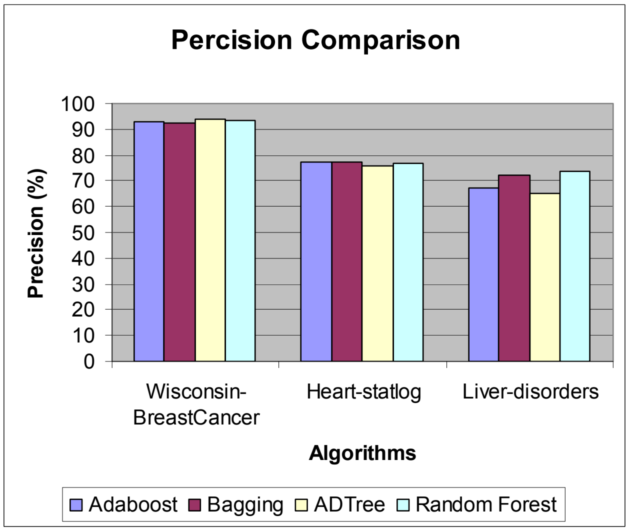


Figure 2: Graphical Representation of Precision

Table 4 is the Error rate comparison of the built models. The graphical version of Error rate comparison is shown in Figure 3.

Table 4: Error Rate Comparison

Dataset	Error Rate (%)			
	Meta Classifier		Decision Tree	
	AdaBoost	Bagging	ADTree	Random Forest
Wisconsin-BreastCancer	5.15	4.43	4.15	3.86
Heart-statlog	20	21.11	21.48	21.85
Liver-disorders	33.91	28.7	40.29	31.01

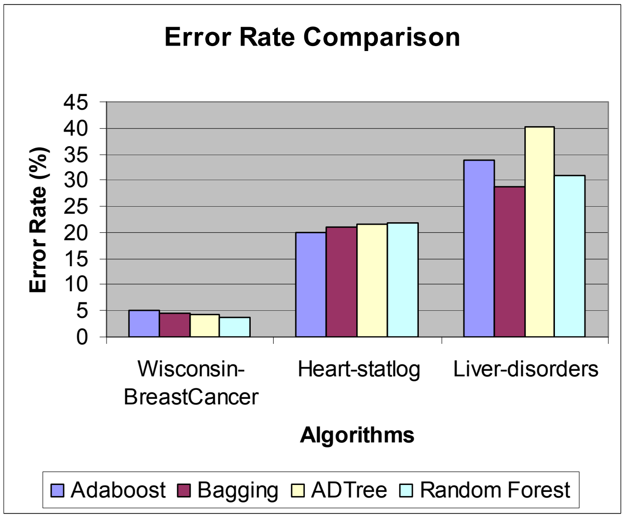


Figure 3: Graphical Representation of Error Rate

Table 5 gives the processing time taken by the algorithms to build the models and the graphical format of execution time comparison is shown in Figure 4.

Table 5: Time taken to build the model

Dataset	**Processing Time (sec)**			
	Meta Classifier		**Decision Tree**	
	AdaBoost	**Bagging**	**ADTree**	**Random Forest**
Wisconsin-BreastCancer	0.3	.45	.33	.55
Heart-statlog	.09	.13	.19	.11
Liver-disorders	.08	.45	.11	.13

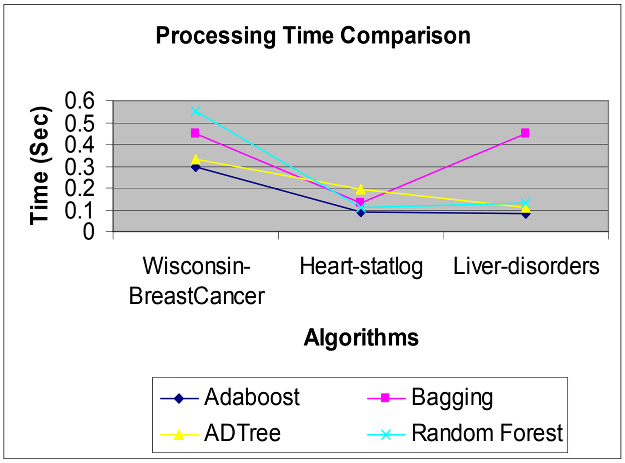


Figure 4: Graphical Representation of Processing Time

V. CONCLUSIONS

In this paper we made an analysis of the accuracy, precision, error rate and the processing time of three medical datasets with different number of instances and number of attributes. The experimental results show that, with the accuracy point of view, the Random Forest works very well in Wisconsin-Breast-Cancer dataset, AdaBoost works better in Heart-statloag and Bagging algorithm gives good result in Liver-disorder dataset. Whereas in precision comparison of the learned model from the available data, the ADTree performs pretty well in Wisconsin-Breast Cancer dataset, and the Random Forest algorithm gives good results in Heart-statlog and Liver-disorders. To be competitive and feasible, it is important to consider the processing time. In our experiments, AdaBoost meta classifier runs in reasonable time in all the three medical datasets. We conclude incisively as a summary for the experimental comparison of bagging and boosting algorithms, No single algorithm performed well for all cases. As the algorithms depends more on dataset than any other factors, a hybrid scheme might be able to combine the advantages of several different approaches. In future, we will perform experimental analysis in combining boosting and bagging techniques in order to build an efficient model with better performance.

VI. REFERENCES

[1] Agarwal. R, Imielinski. T, Swami. A, "Database Mining: A performance perspective", IEEE Transactions on Knowledge and Data Engineering, pp 914-925, December 1993

[2] Anderson, B., & Moore, A. (1998). "Ad-trees for fast counting and for fast learning of association rules". Knowledge Discovery from Databases Conference.

[3] Arlot, S. (2008b). "V -fold cross-validation improved: V -fold penalization". arXiv:0802.0566v2.

[4] Bengio, Y. and Grandvalet, Y. (2004)." No unbiased estimator of the variance of K-fold cross-validation". J. Mach. Learn. Res., 5:1089–1105 (electronic) MR2248010

[5] Bartlett, P. L., & Traskin, M. (2007). "AdaBoost is consistent" Journal of Machine Learning Research, 8, 2347–2368.

[6] Berry Michael J. A. and Linoff Gorden S.,"Mastering Data Mining", John Wiley & Sons, 2000

[7] Bickel, P. J., Ritov, Y., & Zakai, A. (2006). "Some theory for generalized boosting algorithms" Journal of Machine Learning Research, 7, 705–732.

[8] Breiman L, Random Forests, "Machnie Learning", 2001 45(1) pp 5-32

[9] Bouckaert R.R., "Choosing between two learning algorithms based on calibrated tests". In Proceedings of 20th International Conference on Machine Learning. 2003, pp. 51–58.

[10] Chen, M. Han. J. Yu P.S., "Data Mining: An overview from Database Perspective", IEEE Transactions on Knowledge and Data Engineering, Vol 8, No. 6, December 1996.

[11] Collins, M., Schapire, R. E., & Singer, Y. (2002). "Logistic regression, AdaBoost and Bregman distances". Machine Learning, 48.

[12] David Mease, and Abraham Wyner (2008) "Evidence Contrary to the Statistical View of Boosting", Journal of Machine Learning Research 9 131-156

[13] D. Mease, A. Wyner, and A. Buja. "Boosted classification trees and class probability/quantile estimation", Journal of Machine Learning Research, 8:409–439, 2007.

[14] Duda, R. O., Hart, P. E. and Stork, D. G., "Pattern Classification", 2nd Edition, John Wiley & Sons (Asia) PV. Ltd., 2002.

[15] Eric Bauer, Ron Kohavi, "An Empirical Comparison of Voting Classication algorithms: Bagging, Boosting, and Variants", Machine Learning, vv, 1-38 (1998)

[16] Efron, B. and Tibshirani, R. (1997). "Improvements on cross-validation: the .632+ bootstrap method". J. Amer. Statist. Assoc., 92(438):548–560. MR1467848

[17] Han, J., and Kamber, M.,"Data Mining: Concepts and Techniques", 1st Edition, Harcourt India Private Limited. 2001.

[18] Harris Drucker and Corinna Cortes. "Boosting decision trees". In Advances in Neural Information Processing Systems 8, 1996.

[19] Harris Drucker, Robert Schapire, and Patrice Simard. "Boosting performance in neural networks". International Journal of Pattern Recognition and Artificial Intelligence, 7(4):705–719, 1993

[20] J.Han and M. Kamber, "Data mining concepts and Techniques", Morgan Kauffman Publishers, USA, 2006

[21] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: "A statistical view of boosting", Annals of Statistics, 28:337–374, 2000.

[22] Kohavi R., "A study of cross-validation and bootstrap for accuracy estimation and model selection". In Proceedings of International Joint Conference on AI. 1995, pp. 1137–1145, URL http:// citeseer.ist.psu.edu/kohavi95study.html.

[23] Komarek, P., & Moore, A. (2000). "A dynamic adaptation of ad-trees for efficient machine learning on largedata sets". International Conference on Machine Learning (ICML) (pp. 495-502)

[24] Leo Breiman. "Bagging predictors". Technical Report 421, Department of Statistics, University of California at Berkeley, 1994.

[25] Molinaro, A. M., Simon, R., and Pfeiffer, R. M. (2005). "Prediction error estimation: a comparison of resampling methods". Bioinformatics, .3307–3301:(15)21

[26] Mrutyunjaya Panda, Manas Ranjan Patra, "Network Intrusion Detection Using Naïve Bayes", International Journal Of Computer Science And Network Security, VOL.7 No.12, December 2007

[27] Nawei Chen · Dorothea Blostein, "A survey of document image classification: problem statement, classifier architecture and performance evaluation", IJDAR (2007) 10:1–16

[28] Nagy, G.: "Twenty years of document image analysis in PAMI", IEEE Tran. Pattern Anal. Mach. Intell. **22**(1), 38–62 (2000)

[29] Onoda, T., R¨atsch, G., & M¨uller, K.-R. (1998). "An asymptotic analysis of AdaBoost in the binary classification case", Proceedings of the 8th International Conference on Artificial Neural Networks (pp. 195–200)

[30] Patterson, D. W., “Introduction to Artificial Intelligence and Expert Systems”, 8th Edition, Prentice-Hall, India, 2000

[31] Quinlan, J. R., ”Induction of Decision Trees”, Machine Learning, 1:1, Boston: Kluwer, Academic Publishers, 1986, 81-106.

[32] Rich Caruana, Alexandru Niculescu-Mizil, "An Empirical Comparison of Supervised Learning Algorithms", Appearing in Proceedings of the 23 rd International Conference on Machine Learning, Pittsburgh, PA, 2006.

[33] Robert E. Schapire and Yoram Singer. "Improved boosting algorithms using confidence-rated predictions". In Proc. 11th Conf. on Computational Learing Theory, pages 80-91. ACM Press, 1998.

[34] S. B. Kotsiantis, p. E. Pintelas, "Combining Bagging and Boosting", International journal of computational intelligence volume 1 number 4 2004 issn:1304-2386

[35] Yoav Freund, "Boosting a weak learning algorithm by majority", Information and Computation, ,285–256:(2)121 .1995

[36] Shalev-Shwartz, S., & Singer, Y. (2008). "On the equivalence of weak learnability and linear separability: New relaxations and efficient boosting algorithms", 21st Annual Conference on Learning Theory.

[37] Stone M., "Cross-validatory choice and assessment of statistical predictions". J. Royal Stat. Soc., 36(2):111–147, 1974.

[38] Teyssier, M., & Koller, D. (2005). "Ordering-based search: A simple and e ective algorithm for learning bayesian networks", Proceedings of the Twenty-first Conference on Uncertainty in AI (UAI)

[39] Thomas g. Dietterich, "An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees:Bagging, Boosting, and Randomization", Kluwer Academic Publishers, Boston, Machine Learning, , 1-22 (1999)

[40] UCI Machine Learning Repository URL: http://archive.ics.uci.edu/ml/datasets, [accessed on October 2011]

[41] Y. Freund, and R. Shapire, "A decision-theoretic generalization of on-line learning and an application to boosting", Proceedings of the Second European Conference on Computational Learning Theory, 1995, pp. 23 - 37.

[42] Yoav Freund and Llew Mason. "The alternating decision tree learning algorithm", In Proc. 16th Int. Conf. on Machine Learning, pages 124-133. Morgan Kaufmann, 1999.

[43] Yoav Freund and Robert E. Schapire. "Experiments with a new boosting algorithm", In Proc. 13th Int. Conf. on Machine Learning, pages 148-156. Morgan Kaufmann, 1996.

AUTHORS PROFILE

Ms. R. KalaiChelvi Chandrahasan is working as an Asst. Professor in AMA International University, Kingdom of Bahrain. Her research interests are in Cloud Computing, Data mining and Semantic Web mining.

Ms.Angeline Christobel is working as an Asst. Professor in AMA International University, Bahrain. She is currently pursuing her research in Karpagam University, Coimbatore, India. Her research interests are in Data mining, Web mining and Neural networks

Ms.Usha Rani Sridhar is working as an Asst. Professor in AMA International University, Bahrain. Her research interests are in Data mining and Software Engineering

Dr. L. Arockiam is working as an Associate Professor in St.Joseph’s College, India. He has published 89 research articles in the International / National Conferences and Journals. He has also authored two books: "Success through Soft Skills" and "Research in a Nutshell" His research interests are: Software Measurement, Cloud Computing, Cognitive Aspects in Programming, Web Service, Mobile Networks and Datamining

Developing an e-Learning Multiple Choice Questions Test using Mobile SMS

Ali Hussein Ali Alnooh
Computer Science Department
College of Computer Science and Mathematics
Mosul University
Mosul, Iraq
a_alnooh@yahoo.com

***Abstract*—this paper presents a new system for Multiple Choice Questions Paper Test using Mobile SMS (MCQPSMS) to develop the traditional way of MCQ used in Paper Based Tests PBT through the use of mobile Short Message Service (SMS). This MCQPSMS system consists of two main parts: The first one permits the teacher to enter questions and their answers, order them in a random way, then print and give them to students. While the second part receives the answers from the students' mobile phones by SMS, grading them automatically and save them in the database, then sending the marks to the students by SMS. The system has been tested in Mosul university/computer science department with 40 students as a testing sample and the results matched the paper.**

Keywords- MCQ test, E-learning, mobile SMS, AT commands

I. INTRODUCTION

The massive developments in mobile communication systems with the multiplicity of services provided by these systems - particularly SMS - has pushed this development to the need for employing this service in the e-Learning field especially students tests like MCQ test. There are three types of MCQ tests: paper based test, computer based test and mobile based test.

- Paper Based Test (PBT): is easy to implement, traditional and used frequently, but it needs a lot of time from the teacher to grade the answers of the students, also the possibility of teacher's error in answers grading may arise. Some solutions like Object Character Recognition (OCR) systems were adopted to solve those problems, but this need fast scanner devices with high papers per minute (ppm) which cost very high. Some popular organizations use this method like TOEFL or IELETS examination centers.
- Computer Based Test (CBT): depends on web pages and Internet connectivity, meaning that each student must have a computer connected to the Internet during the test time. The possibility of errors during the grading of answers will not arise, but if the Internet connection broke down or the power supply is truncated during the test time, this will lead to restart the test from the beginning. Many popular companies use this kind of test like Microsoft, CISCO, ICDL, etc.
- Mobile Based Test (MBT): this kind of test is very popular in now days because of the wide spread of mobile devices. The test is based mainly on questions supported by multimedia forms like pictures, texts, and voices. The main disadvantage is that the student's mobile device most support multimedia technology used by the test, so if any student doesn't have a suitable mobile cannot participate in the test.

So this paper took the advantages of PBT and MBT by suggesting a new method for testing through PBT, sending the answers using mobile SMS. The test uses both papers and mobile devices, in which the student will test using papers, send the answers in SMS message to the server, the server will grade the answers, save the mark and finally send the result back to the student by SMS.

II. RELATED WORKS

Some related works used the web as a tool for achieving the quizzes without the use of mobile devices. Rarh V. and Goel A. [1] suggested an e-Quizzes system in an interactive manner using the Moodle system, in other word each student must have a computer device connected to the Internet. Other related works employed the mobile phone in the e-Learning environments without using SMS. Tabata Y. and others [2] designed an iphone quiz system for learning the foreign languages by installing this system over students' mobile phones, so if any student doesn't have a device like the iphone cannot attend the exam. Lee K. [3] developed a mobile collaborative learning system through the communication between the students' mobile devices like PDA, phones and pads. Saran M. and others [4] built an e-learning quiz system depending on Multimedia Messaging Service MMS messages and SMS, which means if the student's mobile doesn't support MMS, this will deny the student from the exam.

Also there are other works employed the mobile SMS in controlling and management fields, Givehki F. and Nicknafs A. [5] employed the SMS services to administrate a remote network using the Simple Network Management Protocol SNMP.

III. Short Message Service (SMS) and AT commands

A- SMS: have taken the mobile world by a storm. According researches , there are two types of mobile users: texters and talkers. The texters send more than double the messages that talkers do [6].

So, SMS is a technology that enables the sending and receiving of messages between mobile phones. It was first appeared in Europe in 1992.

The data that can be held by SMS message is very limited. One SMS message can contain at most 140 bytes of data, so one SMS can contain up to 160 characters if 7-bit character encoding is used (like English, Deutsch) or 70 characters if 16-bit character encoding is used (like Arabic, Chinese) [7].

There are many different kinds of SMS applications on the market today and many others are being developed like person-to-person text messaging, provision of information, alerts and notifications.

B- AT Commands:

AT commands are instructions used to control a modem, GSM/GPRS modem or mobile phones. AT is the abbreviation of ATtention. Every command line starts with "AT" or "at".

The AT commands can be executed either by programming or using the Hyper Terminal program.

The starting "AT" is the prefix that informs the mobile phone about the start of a command line. It is not part of the AT command name.

Here are some of the tasks that can be done using AT commands with a GSM/GPRS modem or mobile phone:

- Get basic information about the mobile phone or GSM/GPRS modem.
- Get basic information about the subscriber.
- Send, read, write or delete SMS messages

There are basically two modes to work with SMS: Protocol Data Unit mode (PDU) and Text Mode. A mobile phone internally uses PDU format.

Developers normally uses text mode because it is easier to use. (AT+CMGF) is the command to set the mode, e.g. AT+CMGF=0 sets the PDU mode while AT+CMGF=1 will set the format to text mode. The General syntax of Extended AT commands are:

- All command lines must start with "AT" and end with carriage return character.

- A command line can contain more than one AT commands. Only the first AT command should be prefixed with "AT". AT commands in the same command-line string should be separated with semicolons, e.g. AT+CMGL;+CMGI<CR>.

- A string is enclosed between double quotes, e.g. AT+CMGL="ALL"<CR>.

- Information responses and result codes (including both final result codes and unsolicited result codes) always start and end with a carriage return character and a linefeed character, e.g. after sending the command line "AT+CGMI<*CR*>" to the mobile device, the mobile device should return a response similar to this [8]:

```
<CR><LF>Nokia<CR><LF>
<CR><LF>OK<CR><LF>
```

IV. PROPOSED METHODOLOGY

MCQPSMS system assumes that there are the following hardware requirements:

- Computer device supported by any data base software like SQL, to save the questions with answers and also to save students' names and mobile phone numbers.
- Mobile phone supporting Global System for Mobile communication (GSM) connected to the computer.
- Printer device.

Figure (1) shows the architecture of the system.

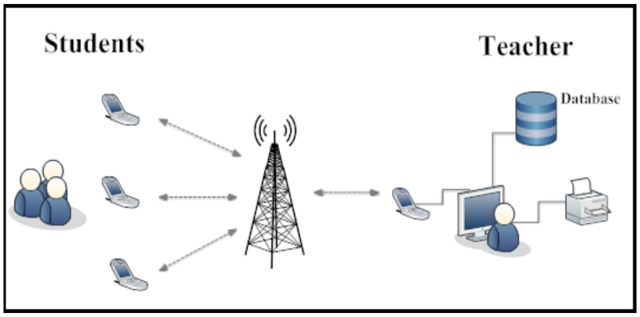


Figure (1) architecture of the system

The teacher will formulate the questions and print them. The students will receive the questions printed on papers, send the answers as SMS to the mobile phone connected to computer.

Now the computer will read the received messages using the AT commands and grade the answers, save and send them to the students. The job is divided into two algorithms, the first one is used by the teacher to formulate and print the questions, while the second algorithm is used to receive the students' answers.

A- First Algorithm:

The teacher can use either saved files containing questions, or write the questions directly with their answers.

The program will generate a random sequence for the questions to each student to ensure that there will be no cheating between students. After that the new sequence will be saved in the DB to be used during marks grading.

Now the program will print the student's name with his/her phone number on the top of the paper and print the questions in the new randomized sequence in the other parts of the paper. Figure (2) shows the flowchart of this algorithm.

B- Second Algorithm:

This algorithm will be implemented after the MCQ test time is ended and all the students sent their answers in SMS messages to the mobile phone connected to the computer beside the teacher.

The program will read all the students' messages and save them in a temporary area, sort them depending on the student's phone and delete the duplicated messages to ensure that if the student has sent more than one answer only the first answer will be depended.

After that the answers will be matched with the randomized sequence saved in DB in algorithm (1) to compute the student's mark.

Now the mark will be saved in the DB and sent in SMS to the students. Each student's result will not take more than 6-10 seconds depending on mobile phone subscriber. Figure (3) shows the flowchart of this algorithm.

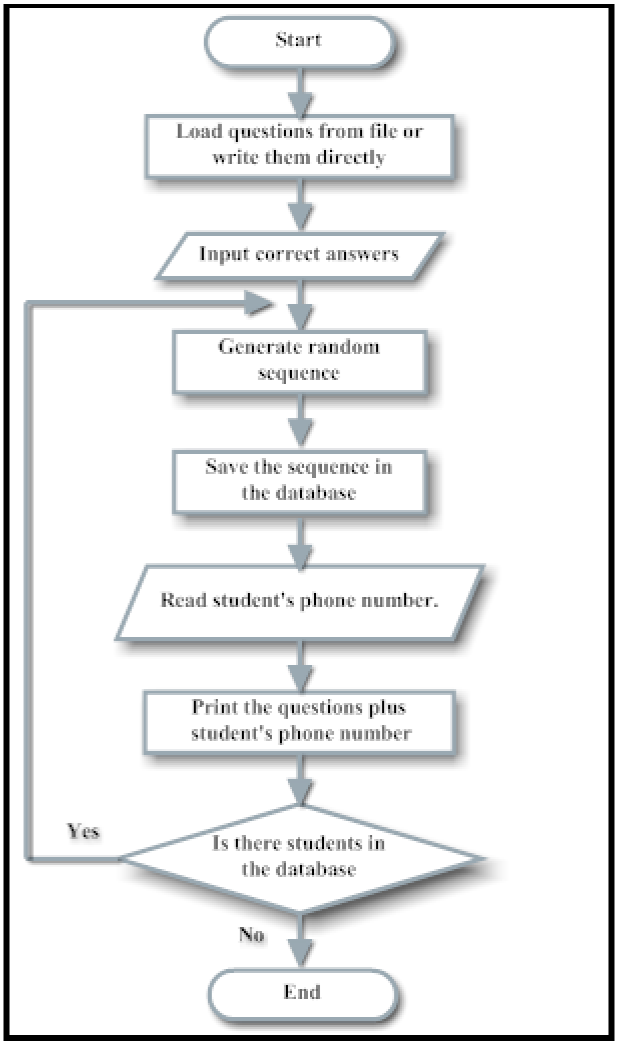

Start
Load questions from file or write them directly
Input correct answers
Generate random sequence
Save the sequence in the database
Read student's phone number.
Print the questions plus student's phone number
Is there students in the database
Yes
No
End

Figure (2) the first algorithm

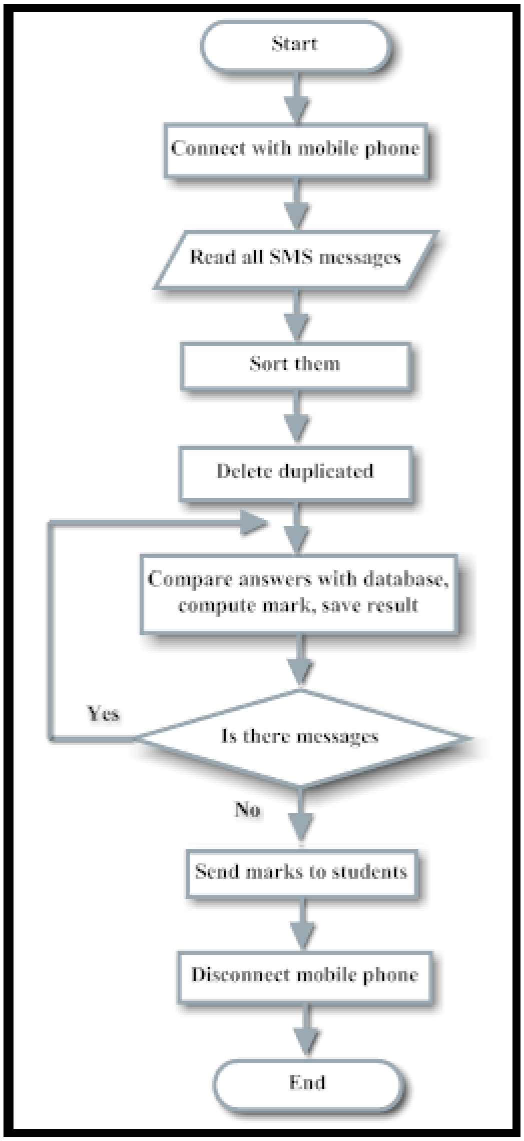


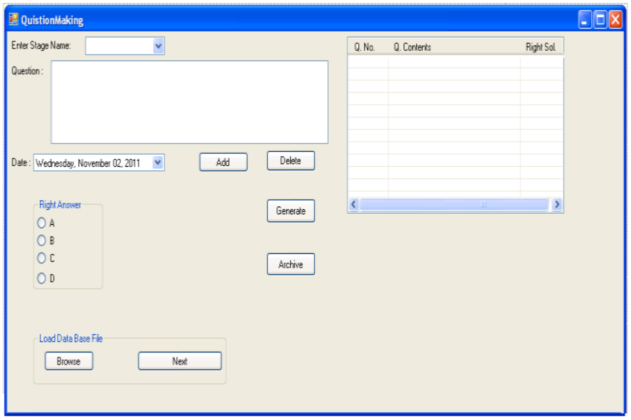
Figure (3) the second algorithm

V. EXPERMINTES AND DISCUSSION

C# language with .NET4 platform was used to program MCQPSMS system. Nokia 6230 was used as a mobile device connected with a computer to receive students' answers.

The system was tested in Mosul University, faculty of computer science and mathematics at the department of computer science with 40 students. The test was composed from 25 MCQ form printed on 4 pages size A4. One of the important features supported by this system is that randomizing the sequence of questions for each student will ensure no one of the students during the exam can send his/her solution to other student.

Figure (4) shows the program interface used by the teacher for formulating and printing the questions. While figure (5) shows the program interface used by the teacher for receiving the SMS messages i.e. answers from the students. Figure (6) shows one of the student's answers.



Date : Wednesday, November 02, 2011

Figure (4) questions formulating and printing

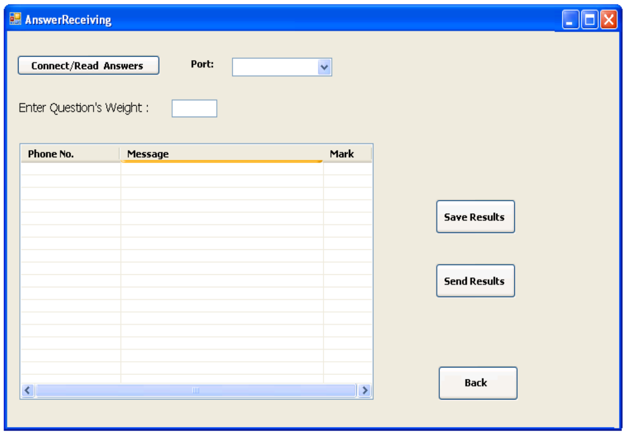
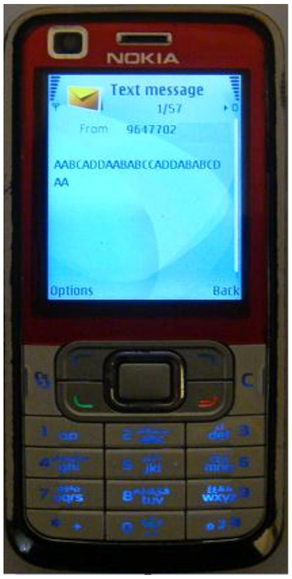


Figure (5) answers receiving



From 9647702

AABCADDAABABCCADDABABCD
AA

Figure (6) sudent's answer

VI SYSTEM EVALUATION

The system was evaluated with the other types of tests, table(1) shows the results of this evaluation:

Factor	PBT	CBT	MBT	MCQPSMS
Communication cost	---	Low	High	Low
Device availability	---	Low	Medium	High
Grading time	Slow	Fast	Medium	Fast
Grading errors	Rarely	Never	Never	Never
Results announcement time	Long	Short	Short	Short

Table (1) system evaluation

For MCQPSMS, the communication cost is low because the cost of the SMS message is very cheap while the cost of MMS - used in MBT - is high.

For the device availabilty, MCQPSMS system is high since all the mobile phone can send SMS messages, while in CBT is low because each student must have a computer connected to Internet during the exam time which is a restricted factor if the number of students exceed the number of available computers. It is medium in MBT because not all of the students have mobile devices supported with multimedia technology.

Fot the grading time plus sending the results, MCQPSMS system is fast since each student does not exceed 6-10 seconds. In PBT this factor is very slow since it is done manually by the teacher, but it can be fast if an OCR and a high speed scanner are used which is costed. In MBT this factor is medium because it depends on MMS .

Also the grading errors in MCQPSMS system, CBT and MBT will never occur because it is done automatically, while it may occur in PBT since it is done manually. Finally, the results' announcements time in MCQPSMS system, CBT and MBT is short, but it takes long time in the PBT.

To evaluate the total time of test grading with results announcement, two tests were done with 25 MCQ printed in 4 pages for three groups consisted from 18,37,40 students respectively.

The tests were done in two ways: **First**, using a PBT with scanner 20 ppm and an OCR software. **Second**, using MCQPSMS system.

For the first group, the average time needed for each student using MCQPSMS is 8 seconds with total time equal to 2.4 minutes, while the average time for each student using OCR is 12 seconds with total time equal to 3.6 seconds, and so on for other groups. The result of the evaluation is shown in figure (7).

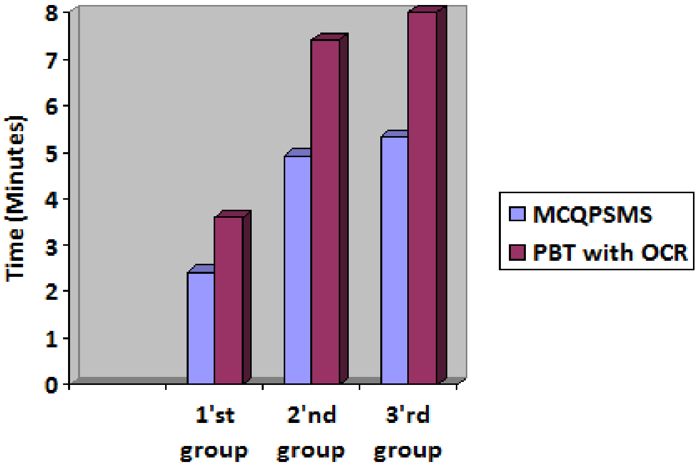


Figure (7) evaluation of grading time

VII CONCLUSIONS

MCQPSMS system can be adopted in educational institutes which relies on PBT, the cost for the requirements is very cheap since it doesn't need more than a traditional mobile phone device and a printer device. Also this system will remove the burdens of manual grading from the teacher, the announcements of the results will be as soon as the test ended.

REFERENCES

[1] Rarh V. and Goel A., "A Methodology for e-Quiz Content Production for E-Learning", 2nd International Conference on Emerging Applications of Information Technology, 2011.

[2] Tabata Y., Yin C., Ogata H. and Yano Y., "An iPhone Quiz System for Learning Foreign Languages", 2nd International Asia Conference on Informatics in Control, Automation and Robotics, 2010.

[3] Lee K. B., "Developing Mobile Collaborative Learning Applications for Mobile Users", International Journal of Interactive Mobile Technologies, Vol 5, No 4, 2011.

[4] Saran N., Kagilaty K. and Saferoglu G., "Use of Mobile Phones in Language Learning: Developing Effective Instuctional Materials", 5th IEEE International Conference on Wireless, Mobile, and Ubiquitous Technology in Education., 2008, pp. 39-43.

[5] Givehki F. and Nicknafs A., "Mobile Control and Management of Computer Networks Using SMS Services", Journal of Telematics and Informatics, Volume 27, Issue 3, 2010.

[6] Mehta N., "Mobile Web Development", Packt publishing, 2008.

[7] Bodic G., "Mobile Messaging Technologies and Services SMS, EMS and MMS", 2nd Ed., John Wiely and Sons, Ltd. 2005.

[8] "AT Commands Reference Guide", Telit Communication Center, 2010.

AUTHORS PROFILE

Mr **Ali H. Alnooh (MSc.)** is currently an assistant lecturer at Mosul University/ College of Coputer Science and Mathematics/Computer Science Department. He obtained his MSc. degree in computer science since 2005, intrested in the area of network security, routing protocols, Mobile networking, e-Learning, e-Government and e-Commerece applications. He teachs Internet Architecture, Computer Networks and Web design for undergraduate students.

DCMC: Decentralized and Cellular Mechanism for improving fault management in Clustered wireless sensor networks

Shahram Babaie[1], Tahereh Rasi[2]
Technical and Engineering Department, Tabriz Branch, Islamic Azad University, Tabriz, Iran
[1]hw.tab.au@gmail.com
[2]tahereh_rasi@yahoo.com

Abstract—Due to the shared wireless communication medium and harsh environments in which sensor nodes are deployed, Wireless Sensor networks (WSN) are inherently fault-prone. Energy is one of the most constraining factors and node failures due to crash and energy exhaustion are commonplace. In order to avoid degradation of service due to faults, it is necessary for the WSN to be able to detect faults early and initiate recovery actions. In this paper we propose a cellular and decentralized cluster based method for any fault detection and recovery which is energy efficient namely DCMC. Simulation Results show that the performance of proposed algorithm is more efficient than previous ones.

***Keywords-** wireless sensor network; fault management; cellular mechanism; cluster-based; energy efficiency*

I. INTRODUCTION

In the recent years, the rapid advances in micro-electro-mechanical systems, low power and highly integrated digital electronics, small scale energy supplies, tiny microprocessors, and low power radio technologies have created low power, low cost and multifunctional wireless sensor devices, which can observe and react to changes in physical phenomena of their environments. These sensor devices are equipped with a small battery, a tiny microprocessor, a radio transceiver, and a set of transducers that used to gathering information that report the changes in the environment of the sensor node. The emergence of these low cost and small size wireless sensor devices has motivated intensive research in the last decade addressing the potential of collaboration among sensors in data gathering and processing, which led to the creation of Wireless Sensor Networks (WSNs).

A typical WSN consists of a number of sensor devices that collaborate with each other to accomplish a common task (e.g. environment monitoring, target tracking, etc) and report the collected data through wireless interface to a base station or sink node. The areas of applications of WSNs vary from civil, healthcare and environmental to military. Examples of applications include target tracking in battlefields [1], habitat monitoring [2], civil structure monitoring [3], forest fire detection [4], and factory maintenance [5].

Due to the deployment of a large number of sensor nodes in uncontrolled or even harsh or hostile environments, it is not uncommon for the sensor nodes to become faulty and unreliable. Fault is an incorrect state of hardware or a program as a consequence of a failure of a component [6]. Some of the faults result from systems or communication hardware failure and the fault state is continuous in time. For example, a node may die due to battery depletion. In this paper we consider only permanent faults, faults occurring due to battery depletion in particular, which when left unnoticed would cause loss in connectivity and coverage.

Faults occurring due to energy depletion are continuous and as the time progresses these faults may increase, resulting in a non-uniform network topology. This often results in scenarios where a certain segment of the network becomes energy constrained before the remaining network. The problems that can occur due to sensor node failure are loss in connectivity, delay due to the loss in connection and partitioning of the network due to the gap created by the failed sensors.

Therefore, to overcome sensor node failure and to guarantee system reliability, faulty nodes should be detected and appropriate measures to recover connectivity must be taken to accommodate for the faulty node. Also, the power supply on each sensor node is limited, and frequent replacement of the batteries is often not practical due to the large number of the nodes in the network. In this paper, we propose a cluster based fault management scheme which detects and rectifies the problems that arise out of energy depletion in nodes. When a sensor node fails, the connectivity is still maintained by reorganization of the cluster. Clustering algorithms such as LEACH [7] and HEED [8] saves energy and reduces network contention by enabling locality of communication.

The localized fault detection method has been found to be energy-efficient in comparison with another algorithm proposed in [9]. Crash faults identification (CFI) [9] performs fault detection for the sensor network. It does not propose any method for fault recovery.

In this paper we propose a cellular approach and decentralized cluster based method called DCMC for fault detection and recovery which is energy efficient.

Corresponding Author: Tahereh Rasi, Technical and Engineering Department, Tabriz Branch, Islamic Azad University, Tabriz, Iran. Email: tahereh_rasi@yahoo.com

The rest of the paper organized as follows: in section 2, we explain the related works. Section 3 describes the proposed algorithm with detailed. Section 4 explore the simulation parameters and result analysis. Final section is containing of conclusion and future works.

II. ELATED WORKS

In this section, we briefly review the related work in the area of fault detection and recovery in wireless sensor networks. Many techniques have been proposed for fault detection, fault tolerance and repair in sensor networks [9, 10, 11, 12]. Cluster based approach for fault detection and repair has also been dealt by researchers in [12]. Hybrid sensor networks make use of mobile sensor nodes to detect and recover from faults [13, 14, 15].

In [16], a failure detection scheme using management architecture for WSNs called MANNA, is proposed and evaluated. It has the global vision of the network and can perform complex tasks that would not be possible inside the network. However, this approach requires an external manager to perform the centralized diagnosis and the communication between nodes and the manager is too expensive for WSNs. Several localized threshold based decision schemes were proposed by Iyengar [11] to detect both faulty sensors and event regions. In [10], a faulty sensor identification algorithm is developed and analyzed. The algorithm is purely localized and requires low computational overhead; it can be easily scaled to large sensor networks. It deals with faulty sensor readings that the sensors report.

In [17], a distributed fault-tolerant mechanism called CMATO for sensor-nets is proposed. It views the cluster as an individual whole and utilizes the monitoring of each other within the cluster to detect and recover from the faults in a quick and energy-efficient way. In fault recovery scheme of this algorithm the nodes within the cluster which its cluster head is faulty join to the neighbor cluster heads which is closest to them.

There have been several research efforts on fault repair in sensor networks. In [18], the authors proposed sensor deployment protocol which moves sensor to provide an initial coverage. In [19], the authors proposed an algorithm called Coverage Fidelity maintenance algorithm (Co-Fi), which uses mobility of sensor nodes to repair the coverage loss. In used solution in [20], the network is partitioned into a virtual grid of cells to perform fault detection and recovery locally with minimum energy consumption. We will refer to this algorithm, with the cellular approach.

To repair a faulty sensor, the work in [14] proposes an algorithm to locate the closest redundant sensor, and use the cascaded movement to relocate the redundant sensor. In [15], the authors proposed a policy-based framework for fault repair in sensor network, and proposed a centralized algorithm for faulty sensor replacement. These techniques outline the ways by which mobile robots/sensors move to replace the faulty nodes. However, movement of the sensor nodes is by itself energy consuming and also to move to an exact place to replace the faulty node and establish connectivity is tedious and energy consuming.

III. PROPOSED PROTOCOL

Due to the large impact of the permanent faults in the cluster head side, in this paper we explore the fault-tolerant mechanism for it.

In this section, we explain the components which are considered in proposed algorithm with details.

A. Network Model

Let us consider a sensor network which consists of N nodes uniformly deployed over a square area with high densely. There is a sink node located in the field, and the cluster heads use multi-hop routing to send data to it. Also the nodes in each cluster use tree topology to send data to cluster head. We assume all nodes, including the cluster heads and the normal nodes, are homogeneous and have the same capabilities, and they use power control to vary the amount of transmission power which depends on the distance to the receiver.

As can be seen in Fig. 1, this algorithm selects a node as a manager node in each cluster so that firstly it has maximum remained energy and secondly it has maximum number of ordinary nodes in its neighborhood. For this reason, this algorithm uses (1) to select cluster manager.

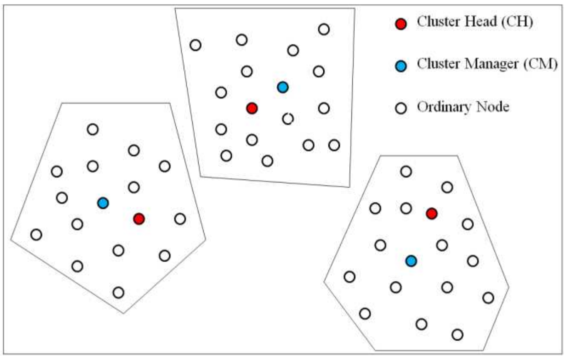


Figure 1. Network model in DCMC

$$M\_V_{CM} = \alpha(\frac{E_r}{E_i}) + \beta(\frac{N_{non}}{N_{on}}) \qquad (1)$$

Here, E_r is the remaining energy of the node and E_m is the amount of its initial energy. N_{non} of a node is the number of neighboring ordinary nodes which is in its transmission radio range and N_{on} is the number of all ordinary nodes in the cluster. Parameters α and β determines the weight of each ratio so that sum of them is 1.

The node that has higher merit value is selected as a cluster manager and hence it is responsible for fault management in the cluster. In fact, the node has the merit of being cluster manager that amount of its energy is more and it is also located in center of the cluster.

Then, the node that is selected as a cluster manager divides the cluster into four cells virtually so that it places in center of cells (as can be seen in Fig. 2). In addition, the cluster manager selects a cell manager for each cell by (2).

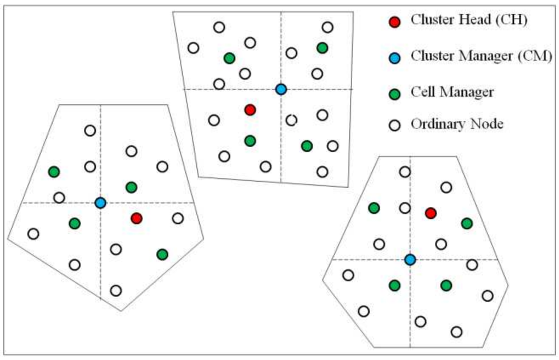


Figure 2. Dividing the cluster into four virtuall cells by cluster manager

$$M\_V_{CeM} = \alpha(\frac{E_r}{E_i}) + \beta(\frac{N_{nonc}}{N_{onc}}) + \lambda(\frac{\Sigma E_{r\_nnonc}}{\Sigma E_{i\_nnonc}}) \tag{2}$$

In here, N_{nonc} of a node is the number of neighboring ordinary nodes which is in its transmission radio range in the same cell and N_{onc} is the number of all ordinary nodes in the cell. $E_{r\text{-}nnonc}$ is remaining energy of non-neighboring ordinary node and $E_{i\text{-}nnonc}$ is its initial energy. Parameters α, β and λ determine the weight of each ratio so that sum of them is 1.

The third parameter indicates the amount of remaining energy of non-neighboring nodes should be more. As can be inferred, the possibility of failing non-neighboring nodes due to the high remaining energy of these nodes is low. Therefore, the energy consumed by these nodes and the cell manager to the fault management in the cell decreases.

Cluster manager is connected only to cluster head and cell managers. Cell managers are connected only to cluster head, cluster manager and nodes within the cell in proposed protocol. We assume that these connections are single hop.

B. Energy Consumption Model

In DCMC, energy model is obtained from [7] that use both of the open space (energy dissipation d^2) and multi path (energy dissipation d^4) channels by taking amount the distance between the transmitter and receiver. So energy consumption for transmitting a packet of l bits in distance d is given by (3).

$$E_{Tx}(l,d) = \begin{cases} lE_{elec} + l\varepsilon_{fs} d^2 & , d \le d_0 \\ lE_{elec} + l\varepsilon_{mp} d^4 & , d > d_0 \end{cases} \tag{3}$$

Here d_0 is the distance threshold value which is obtained by (4), E_{elec} is required energy for activating the electronic circuits. ε_{fs} and ε_{mp} are required energy for amplification of transmitted signals to transmit a one bit in open space and multi path models, respectively.

$$d_0 = \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}} \tag{4}$$

Energy consumption to receive a packet of l bits is calculated according to (5).

$$E_{Rx}(l) = lE_{elec} \tag{5}$$

C. Fault Detection and Recovery

Managers of the cells, cluster manager and cluster head in the cluster cooperate with each other to detect the faulty node in their own cluster. Each of these nodes cooperate in detect the failure of different types of nodes in different forms. These operations discuss in below.

1) Fault Detection and Recovery in the Ordinary Nodes

Initially, the managers of all cells are in sleep mode. When cluster head does not receive the data in response the sent *Data_Req* packet, it will immediately inform the cluster manager. Then cluster manager send an *Awake* message to manager of the cell that desired node is located on it. Then the cell manager sends a Query message to the desired node and requests from it to responds to cell manager and cluster manager. Finally, the majority voting is done between cell manager, cluster manager and cluster head to determine the status of desired node. The desired node is not used in the next, if it is detected failed. Fig. 3 shows these operations.

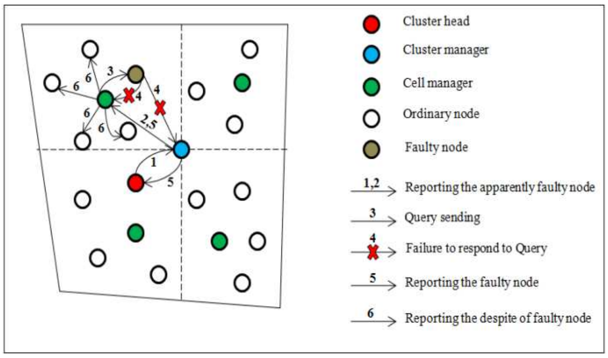

6
3
6
4
4
6
2,5
6
1
5
Cluster head
Cluster manager
Cell manager
Ordinary node
Faulty node
1,2 Reporting the apparently faulty node
3 Query sending
4 Failure to respond to Query
5 Reporting the faulty node
6 Reporting the despite of faulty node

Figure 3. Operations of fault detection and recovery in the ordinary nodes

2) Fault Detection and Recovery in the Cluster Head

For this purpose, cluster manager sends the Query message to cluster head and gets its response message periodically. Cluster manager reports to the cell managers when it does not receive any response from the cluster head. Cell managers also repeat this operation. Then the majority voting is done between cell managers and cluster manager to determine the status of desired cluster head.

Cluster manager will select a new cluster head from among the nodes by (6) if the current cluster head is detected failed.

$$M\_V_{New\_CH} = \frac{E_r}{(D_{nch\_och})^2} \tag{6}$$

Here, E_r is remaining energy of the node which will be chosen as the new cluster head. $D_{nch\text{-}och}$ is the distance between faulty cluster head and the node which will be chosen as the new cluster head.

3) *Fault Detection and Recovery in Cluster Manager*

Early detection of failure in the cluster manager is done by the cluster head. So that, if the cluster head does not receive any Query message from cluster manager in a specified time, it detects that the cluster manager is probably failed. Then it will report this status to the cell managers. Then the managers of all cells and cluster head will send the Query message to the cluster manager to determine its status. Then the majority voting is done between cell managers and cluster head to determine the status of cluster manager.

Cluster head informs all nodes to compete with each other for selection as a new cluster manager if the cluster manager is detected failed.

4) *Fault Detection and Recovery in the Cell Manager*

Early detection of failure in the cell manager is done by the cluster manager. So that, cluster manager determines that the cell manager is failed when it does not receives any response from cell manager for sent Awake message. In this case, cluster manager informs the two cell managers that are adjacent to the desired cell manager. Then these neighboring cell managers send a Query message to desired cell manager to determine its status. Then the majority voting is done between these two cell managers and cluster manager to determine the status of desired cell manager. Fig. 4 shows these operations.

Cluster manager informs all nodes of desired cell to compete with each other for selection as a new cell manager if desired cell manager is detected failed.

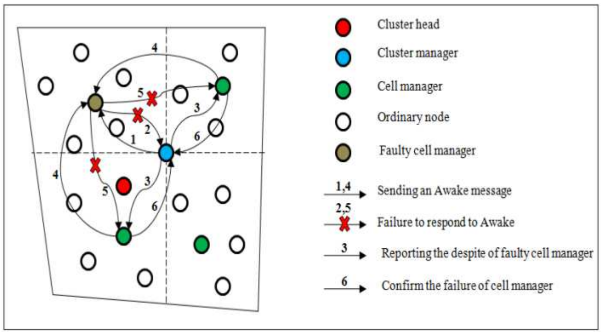

Cluster head
Cluster manager
Cell manager
Ordinary node
Faulty cell manager
1,4 Sending an Awake message
2,5 Failure to respond to Awake
3 Reporting the despite of faulty cell manager
6 Confirm the failure of cell manager
1
2
3
4
5
6

Figure 4. Operations of fault detection and recovery in the cell manager

IV. Simulation and Performance Evaluation

In this section, we present and discuss the simulation results for the performance study of DCMC protocol. We used GCC to implement and simulate DCMC and compare it with the Cellular approach protocol.

The network is clustered using the HEED clustering algorithm, the cluster heads then organize into a spanning tree for routing. We implement DCMC on HEED protocol. Simulation parameters are presented in Table I and obtained results are shown below.

TABLE I. Simulation Parameters

Parameters	Value
Network area	400 meters × 400 meters
Sink location	(0, 0)m
Number of sensors	100
Initial energy	2J
E_{elec}	50 nJ/bit
ε_{fs}	10 pJ/bit/m^2
ε_{mp}	0.0013 pJ/bit/m^4
d_0	87 m
E_{DA}	5 nJ/bit/signal
Query packet size	20 bytes
Data packet size	512 bytes

Fig. 5 shows the average energy loss for fault detection in DCMC and Cellular approach. In this evaluation, we change the transmission range at the all nodes from 20 meters to 60 meters and calculate the energy loss for fault detection.

As it can be seen, proposed protocol has performance better than Cellular approach in average energy loss for fault detection.

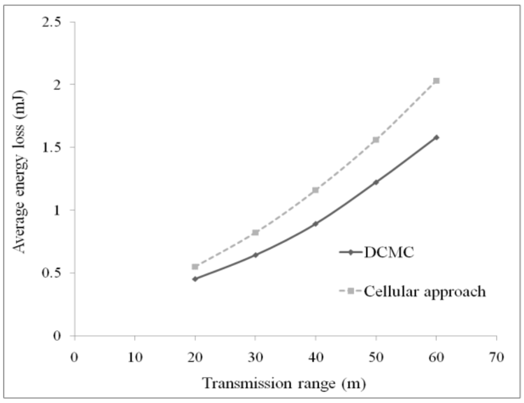


Figure 5. Average energy loss for fault detection

Fig. 6 shows the fault diagnosis accuracy for DCMC and Cellular approach. In this evaluation, we set the link failure rate of some links between 10% and 30% randomly and measure the accuracy of both protocols on the detection of failure.

It can be observed that the proposed method has performance better than Cellular approach in fault diagnosis accuracy. This is because that the fault detection algorithm in DCMC uses the majority voting mechanism.

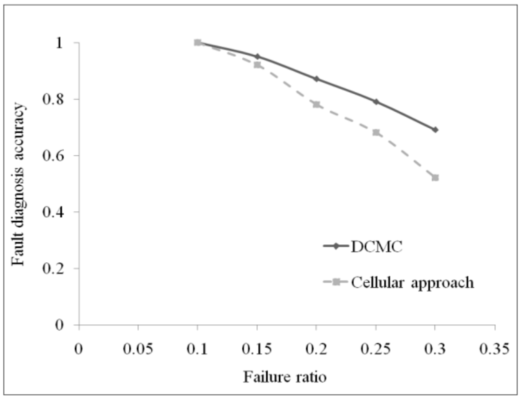


Figure 6. Fault diagnosis accuracy

I. Conclusion

In this paper we propose a cellular and decentralized cluster based method for any fault detection and recovery which is energy efficient namely DCMC. Simulation Results show that the DCMC consumes less energy for fault detection and recovery in comparison to Cellular approach. In addition, DCMC has more accuracy in fault diagnosis. In general, performance of DCMC is more efficient than Cellular approach.

References

[1] T. Bokareva, W. Hu, S. Kanhere, B. Ristic, N. Gordon, T. Bessell, M. Rutten and S. Jha, "Wireless Sensor Networks for Battlefield Surveillance", In roceedings of The Land Warfare Conference (LWC)– October 24 – 27, 2006, Brisbane, Australia.

[2] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Anderson, “Wireless Sensor Networks for Habitat Monitoring,” in the Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications (ACM-WSNA), Pages: 88-97, September 28 - 28, 2002, Atlanta, Georgia, USA.

[3] N. Xu, S. Rangwala, K. Chintalapudi, D. Ganesan, A. Broad, R. Govindan, and D. Estrin, “A Wireless Sensor Network for structural Monitoring,” in Proc. ACM SenSys Conf., Nov.2004.

[4] M. Hefeeda, M. Bagheri, “Wireless Sensor Networks for Early Detection of Forest Fires”, in the proceedings of IEEE Internatonal Conference on Mobile Adhoc and Sensor Systems, 2007. MASS 2007. Volume , Issue , 8-11 Oct. 2007 Page(s):1 – 6, Pisa, Italy.

[5] K. Srinivasan, M. Ndoh, H. Nie, H. Xia, K. Kaluri, and D. Ingraham, “Wireless Technologies for Condition-Based Maintenance (CBM) in Petroleum Plants,” Proc. of DCOSS’05 (Poster Session), 2005.

[6] F. Koushanfar, M. Potkonjak, and A. Sangiovanni-Vincentelli, “Fault tolerance in wireless ad hoc sensor networks,” in IEEE Sensors, vol. 2,pp. 1491-1496, June 2002.

[7] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, “Energy-Efficient Communication Protocol for Wireless Microsensor Networks,” Proceedings of the Hawaii International Conference on System Sciences, 2000.

[8] O. Younis, and S. Fahmy, “HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks,” IEEE Transactions on Mobile Computing, vol. 3, no. 4, 2004, pp. 366-379.

[9] S. Chessa and P. Santi, “Crash Faults Identification in Wireless Sensor Networks,” in Computer Comm., vol. 25, no. 14, pp. 1273-1282, Sept. 2002.

[10] M. Ding, D. Chen, K. Xing, X. Cheng, “Localized fault-tolerant event boundary detection in sensor networks,” in IEEE Infocom, March 2005.

[11] B. Krishnamachari and S. Iyengar, “Distributed Bayesian Algorithms for Fault-tolerant Event Region Detection in Wireless Sensor Network, ” in IEEE Transactions on Computers, 53, 3, 241-250, March 2004.

[12] G. Gupta, M. Younis. “Fault-tolerant clustering of wireless sensor networks,” in Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE Volume 3, 16-20 March 2003 Page(s):1579 - 1584 vol.3.

[13] Y. Mei, C.Xian,S.Das,Y.C.Hu and Y.H Lu, “Repairing Sensor Networks Using Mobile Robots,” in Proceedings of the ICDCS International Workshop on Wireless Ad Hoc and Sensor Networks (IEEE WWASN 2006), Lisboa, Portugal, July 4-7, 2006.

[14] G. Wang, G. Cao, T. Porta, and W. Zhang, “Sensor relocation in mobile sensor networks,” in the 24th Conference of the IEEE Communications Society (INFOCOM), March 2005.

[15] T. Le, N. Ahmed, N. Parameswaran, and S. Jha, “Fault repair framework for mobile sensor networks,” in IEEE COMSWARE, 2006.

[16] L. B. Ruiz, I. G. Siqueira, L. B. Oliveira, H. C. Wong, J. M. S.Nogueira, and A. A. F. Loureiro, “Fault management in event-driven wireless sensor networks, ” in MSWiM ’04: Proceedings of the 7th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems, pages 149–156, New York, 2004.

[17] Y. Lai, and H. Chen, “Energy-Efficient Fault-Tolerant Mechanism for Clustered Wireless Sensor Networks”, Proceedings of 16th International Conference on Computer Communications and Networks, pages 272-277, 2007.

[18] G. Wang, G. Cao, and T. L. Porta, “A bidding protocol for deploying mobile sensors,” in 11th IEEE International Conference on Network Protocol ICNP ’03, pp. 315–324, Nov 2003.

[19] S. Ganeriwal, A. Kansal, and M. B. Srivastava, “Self aware actuation for fault repair in sensor networks,” in IEEE International Conference on Robotics and Automation (ICRA), May 2004.

[20] M. Asim, H. Mokhtar, and M. Merabti, “A cellular approach to fault detection and recovery in wireless sensor networks”,in:Third International Conference on Sensor Technologies and Applications, 2009.

Comparative Study of the Effectiveness of Ad Hoc, Checklist- and Perspective-based Software Inspection Reading Techniques

Olalekan S. Akinola
Solom202@yahoo.co.uk
Department of Computer Science,
University of Ibadan, Ibadan, Nigeria

Ipeayeda Funmilola Wumi
funmipy12@yahoo.com
Department of Computer Science,
University of Ibadan, Ibadan, Nigeria

ABSTRACT

Software inspection is said to be inevitable in order to ensure software quality assurance. Nevertheless, there have been controversies on which defect detection techniques should be applied in software document inspection. This work comparatively study the effectiveness of three software inspection techniques: Ad *Hoc*, Perspective-based and Checklist-based defect detection techniques. Paper-based inspections of software artifact were carried out on an industrial code artifact seeded with forty bugs. An experimental 3 x 3 x 4 factorial design with three defect detection techniques (checklist-based, Adhoc and perspective-based) as independent variables, three dependent variables (inspection effectiveness, effort and false positives) and four teams for each defect detection methods was used for the experiment. The data obtained were subjected to tests of hypotheses using One-way ANOVA, Post-Hoc tests and Mean coefficients. Results from the study indicate that there were significant differences in the defect detection effectiveness and effort in terms of time taken in minutes reported by the reviewers using perspective-based, ad hoc and checklist-based based reading techniques in the industrial settings.

Key words: Software inspection, Ad Hoc reading technique, Checklist reading technique, Perspective reading technique

1. INTRODUCTION

The process of improving software quality has been a growing discussion in the few decades. Software quality can be defined to be software that satisfies the needs of the users and the programmers involved in it or as the customer's perception of how the system work. Software inspection is a fundamental component of the software quality assurance process. It is a process whereby a group of software competent people critically checks a piece of software milestone for detecting defects [1]. Inspection improves the quality of software products, such as understand-ability, portability, maintainability, testability, etc. Its success has always been demonstrated in many published articles.

Software Inspections are a formalized, structured form of *peer reviews*. They are an extremely cost-effective quality assurance technique that can be applied to any type of software project deliverable, such as Requirements documents, Design documents, Code, and other items such as test plans and user documents. For most software organizations, Software Inspections are *the most important single process improvement*. According to Capers [4], "*… formal design and code inspections rank as the most effective methods of defect removal yet discovered … (defect removal) can top 85%, about twice those of any form of testing.*"

Since the year Fagan developed the inspection process in the early 70s at IBM; there have been many variations of the process put forth by others. The aim is to uncover faults in the products, rather than to correct them. The goal of inspection meetings is to collect the faults discovered and bring synergy (process gains) to the software inspection. It is believed that the combination of different viewpoints, skills and knowledge from many reviewers creates this synergy [10].

Ad Hoc, checklist-based and perspective-based reading techniques are the three commonly used inspection artifacts reading techniques. To the best of our knowledge, experiments comparing the effectiveness of all three together are scarce. This research is therefore conducted to find if there are any significant differences in the effectiveness of reviewers using Perspective-based, Checklist-based and Ad hoc code reading techniques in an industrial code setting. Thirty volunteered reviewers from ten software houses in Nigeria were used to carry out code inspection on the visual-basic large-sized code artifact.

1.1 Research Hypotheses

Three hypotheses were stated for this experiment as follows.

Ho1: There is no significant difference among the effectiveness of reviewers using Perspective-based, Ad hoc and Checklist reading techniques in distributed code inspection.

Ho2: There is no significant difference among the effort taken by reviewers using Perspective-based, Ad hoc and Checklist techniques in distributed code inspection.

Ho3: There is no significant difference among the false positives reported by reviewers using Perspective-based, Ad hoc and Checklist techniques in distributed code inspection.

2. SOFTWARE INSPECTION READING TECHNIQUES

Software inspection encompasses a set of methods in which the purpose is to identify and locate faults in software. Software inspection is a peer review process led by software developers who are trained in inspection techniques [7]. Michael Fagan [10] originally developed the software inspection process ‘out of sheer frustration’ [13]. Since Fagan developed the inspection process in the early 1970s at IBM, there have been many variations of the process put forth by others. Overall, the aim in any review process is to apply inspection to the working product as early as possible so that major faults are caught before the product is released.

A reading technique can be defined as a series of steps or procedures whose purpose is to guide an inspector in acquiring a deep understanding of the inspected software product [8]. The comprehension of inspected software products is a prerequisite for detecting subtle and/ or complex defects, those often causing the most problems if detected in later life-cycle phases. In a sense, a reading technique can be regarded as a mechanism or strategy for individual inspector to detect defects in the inspected product.

There are many reading techniques that focus on finding as many detects as possible but three among the reading techniques were used in carrying out this work; ad hoc, checklist-based and perspective-based reading techniques. According to Porter and Votta [11], defect detection techniques range in prescription from intuitive, non-systematic procedures such as ad hoc or checklist techniques, to explicit and highly systematic procedures such as Perspective technique.

Ad-hoc reading, by nature, offers very little reading support at all since a software product is simply given to inspectors without any direction or guidelines on how to proceed through it and what to look for. However, ad-hoc does not mean that inspection participants do not scrutinize the inspected product systematically. The word ad-hoc only refers to the fact that no technical support is given to them for the problem of how to detect defects in a software artifact. In this case, defect detection fully depends on the skill, the knowledge, and the experience of an inspector. Training sessions in program comprehension as presented in Rifkin and Deimel [12], may help subjects develop some of these capabilities to alleviate the lack of reading support.

Perspective-based reading technique gives reviewers a set of procedures to inspect software products for defects. The perspective-based reading technique instructs the reviewer to perform an active review by assigning different perspectives to each reviewer. Common perspectives are user, tester, and designer.

Checklist-based reading technique reviewers use a checklist which guides them regarding what kind of faults to look for. The reviewers read the document using the checklist to guide the focus of their review. Checklist-based offers stronger, boilerplate support in the form of questions inspectors are to answer while reading the document

3. RESEARCH METHODOLOGY

3.1 Subjects

The subjects used for this research were the Software professionals drawn from ten software houses in Nigeria. Software professionals were chosen as subjects for this research because results obtained with professionals would make us to predict what may likely happen at industry level.

3.2 Experimental artifact and Instrumentation

The artifact inspected was a large sized Visual basic 6.0 language industrial code. It calculates staff allowances such as domestic, responsibility, hazard, housing, leave, medical, transport, utility and so on in order to compute staff salary monthly, annual salary and arrears as well as some deductions(such as tax, cooperative, official housing accommodation rent and so on) to be made on staff salary. The artifact which was 500 lines of code was tested okay before it was seeded with 40 bugs which are syntax, semantics and logical in nature.

The designed instruments for this experiment were the individual preparation forms, the experimental artifact (code) and the collection meeting forms. The experimental artifact and individual preparation forms were given to each reviewer. The individual preparation form was filled during preparation by each reviewer. During individual preparation, each reviewer recorded the start and the end times for the review of the artifact. The line number of the suspected defect and the description of defects suspected were also recorded in the forms.

The meeting form was filled in at the collection meeting. The start and end times of the team collection meetings held was filled on the collection meeting forms. The line numbers of the defect and the defect description was recorded on the collection meeting forms. Most importantly, the teams' identification numbers were filled on the collection meeting form in order to identify each team.

3.3 Conducting the Experiment

The experiment was monitored and conducted by the researchers. The software professionals were used for the experiment without specifying any particular year of experience. There was no special training given to them on Visual Basic programming because they are software professionals who are conversant with the language used for the artifact. Nevertheless, the reviewers were given some initial code inspection trainings before the real experiments were carried out.

During individual preparation, reviewers examined the artifact in order to identify the bugs seeded in them. In perspective-based technique, reviewers were assigned with a particular role to play (a designer, tester, reader and user) and in checklist-based, a Visual basic checklist question designed by the researchers was given to reviewers in order to guide them in fishing out bugs. There was no particular time given to the reviewers for their individual preparations. All the suspected defects were recorded on the individual preparation form.

Moreover, before the commencement of the team collection meeting, the individual preparation forms were collected by the researcher in order that the reviewers do not add to their preparation forms any defects that were not found during team defects collection meeting. During defects collection meetings, there was no specific duration given to the reviewers for the artifact inspections.

During the defect collection meetings, one of the reviewers in each team serves as the reader, moderator and recorder. While in the meetings, reviewers brought up new defects or discussed any defects found during the individual preparation. All defects found were recorded in the team defects collection meeting forms by the recorder.

Four different teams were created each for the Perspective-based, Ad hoc and Checklist-based reviewers. In order to eliminate bias in the results, team sizes were duplicated for each of the groups. For instance, teams 001A and 001B in Table 1 are for team size of one, while teams 002A and 002B are of size two. This is done for team sizes three and four in that order.

3.4 Threats to Validity of the Experiment

In this experiment, two important threats that may affect validity of the research are considered. These threats limit our ability to generalize or guarantee the results and hence, it demands caution when interpreting the results. In this experiment, we considered two important threats that may affect the validity of the research in the domain of software inspection.

3.4.1 Threats to Internal Validity

Threats to internal validity are influences that can affect the dependent variables without the researcher's knowledge [15]. We considered three such influences: (1) selection effects, (2) maturation effects, and (3) instrumentation effects.

Selection effects are due to natural variation in human performance [14]. We limited this effect by randomly assigning team members for the inspection. This way, individual differences were spread across all treatments.

Maturation effects result from the participants' skills improving with experience. If the same set of participants were used in all three experiments, there may be maturation effect also as the participants' inspection ability may get better over time. Randomly assigning the reviewers and doing the review within the same period of time checked these effects.

Instrumentation effects are caused by the artifacts to be inspected, by differences in the data collection forms, or by other experimental materials. In this study, this was negligible or did not take place at all since all the groups inspected the artifacts within the same period of time. Again, one set of data collection forms was used for all the groups (the treatments).

3.4.2 Threats to External Validity

Threats to external validity are conditions that can limit our ability to generalize the results of experiments to industrial practice [14]. We considered one source of such threats: experimental scale.

Experimental scale is a threat when the experimental setting or the materials are not representative of industrial practice. This study made use of a Visual Basic industrial code that computes wages and salary of staff in a particular company. More so, industrial experienced software professionals were used for the experiment. Therefore, experimental scale effect was reduced to a large extent in this study

4. RESULTS

We are particular about the effectiveness of the reviewers, the effort in terms of number of minutes taken by them and the false positives reported by the reviewers in a defect collection meeting. These three data were collected from the experiment. Initially the reviewers were given the code artifact to study individually at preparations before the actual defect collection meeting took place. Table 1 gives the mean aggregate values of results obtained with the inspection teams.

The teams' effectiveness at detecting errors in the code artifact is depicted in Figure 1. Fp means false positives reported by the reviewers. False positives are those errors perceived by the reviewers to be true errors but were indeed not valid.

Table 1: Raw data of Team collection meetings

	PERSPECTIVE			CHECKLIST-BASED			AD HOC		
Teams	Defects	Effort (Mins)	Fp	Defects	Effort (Mins)	Fp	Defects	Efforts (Mins)	Fp
001A	31	58	6	25	71	4	29	43	3
001B	34	56	4	28	67	5	28	63	3
002A	26	46	8	30	71	6	28	58	5
002B	36	61	5	28	69	5	27	49	3
003A	28	58	5	29	69	5	28	38	3
003B	34	68	4	24	66	2	31	58	5
004A	30	45	6	27	59	4	28	48	6
004B	31	38	5	29	44	3	25	41	4

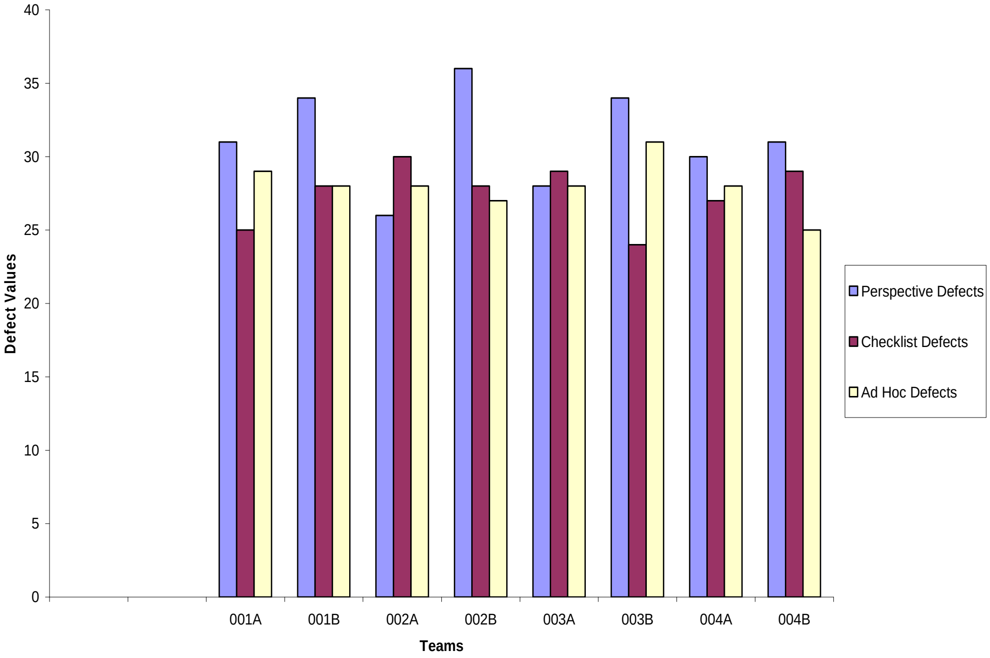


Figure 1: Teams’ effectiveness at detecting errors in the code artifact

Figure 1 shows that Perspective –based teams have the highest defect detection effectiveness compared to the other groups – Checklist-based and Ad hoc groups.

Figure 2 shows that Efforts in terms of minutes expended by the teams were high and roughly the same for most of the Checklist-based teams even though there were no restrictions placed on the time the reviewers must spend for the code inspection.

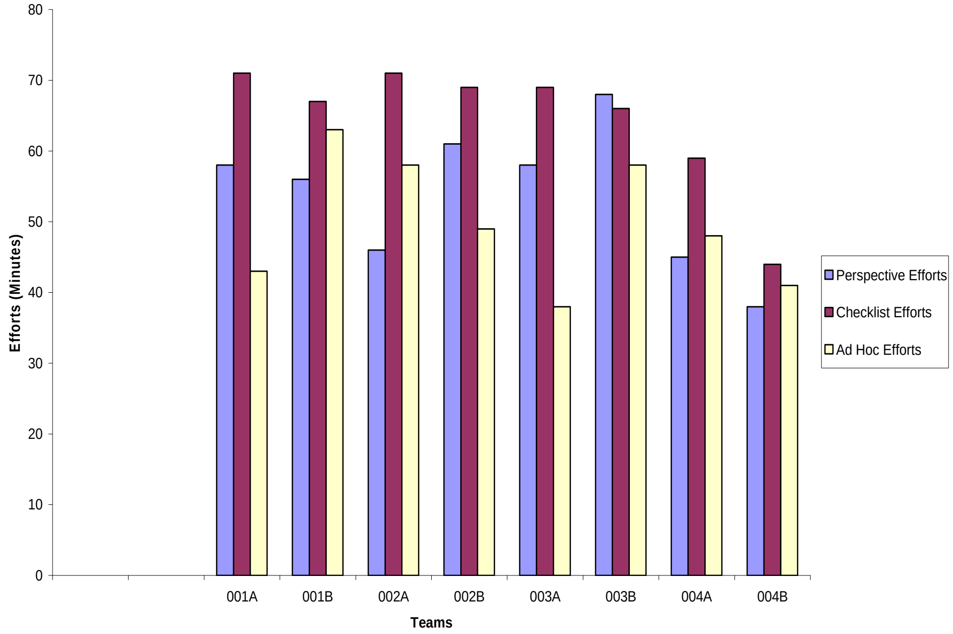


Figure 2: Teams' efforts (Minutes) expended in the code Inspection

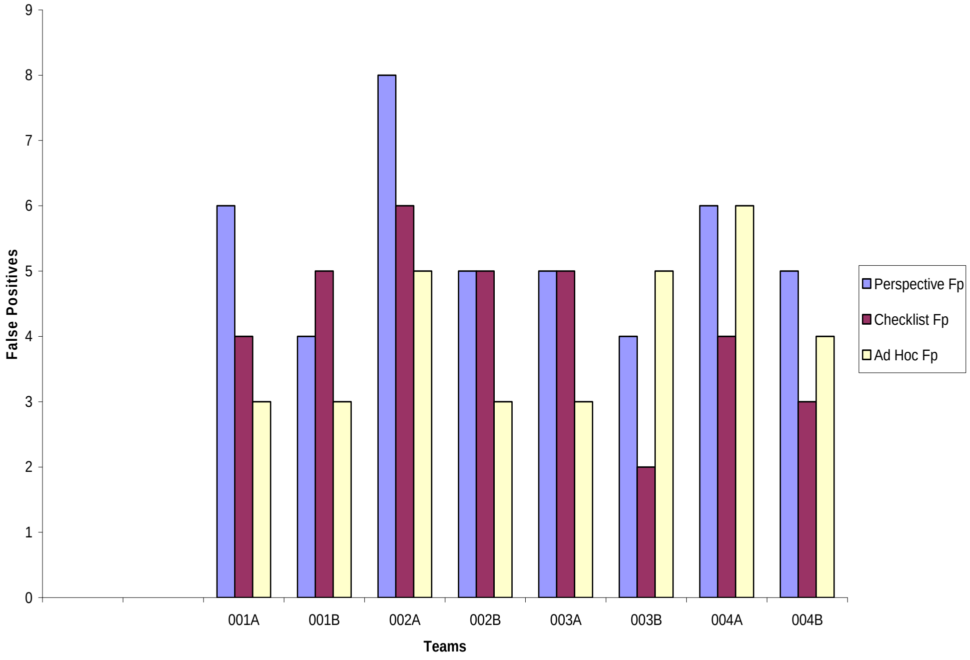


Figure 3: Teams' False positives reported in the code Inspection

Figure 3 shows that perspective-based teams which had highest defect detection effectiveness from Figure 1 reported more or less highest number of false positives, especially teams 001A and 002A.

4.1 Further Statistical Analyses

The data obtained in the experiment were subjected to further statistics tests. One-Way ANOVA was used to perform the statistical analyses and thereafter Post-Hoc Turkey HSD and Duncan tests were done to determine where differences lie among the data pairs.

Table 3 shows the results of ANOVA statistical tests performed on the data obtained in this experiment.

Table 3 shows that there was a significant difference among the defect detection effectiveness of the reviewers using the three reading techniques – Perspective-based, checklist-based and ad hoc. This is also true of the effort (time) expended by the reviewers during the inspection exercise. However, the case is different with False positives reported by the reviewers.

Further Tukey HSD post hoc test analyses was carried out on the data to ascertained where the differences actually lie. The results of the analyses are shown in Table 4.

Table 3: Results of ANOVA Analysis

Data	Hypothesis Tested	p-value	Mean values	Decision
Defect Detection Effectiveness (effe)	Ho: There is no significant difference among the effectiveness of reviewers using Perspective-based, Ad hoc and Checklist-based techniques	0.012	$Persp_{effe}$ = 31.25 CBR_{effe} = 27.50 AH_{effe} = 28.00	H_1 accepted $P < 0.05$
Effort (eff, in Minutes)	Ho: There is no significant difference among the effort of reviewers using Perspective-based, Ad hoc and Checklist-based technique	0.014	$Persp_{eff}$ = 53.75 CBR_{eff} = 64.50 AH_{eff} = 49.75	H_1 accepted $p < 0.05$
False Positives (fp)	Ho: There is no significant difference among the false-positives of reviewers using Perspective-based, Ad hoc and Checklist-based technique	0.090	$Persp_{fp}$ = 5.37 C_{fp} = 4.25 A_{fp} = 4.00	H_0 accepted $p > 0.05$

Table 4. Tukey HSD Post Hoc Multiple Comparison Tests Summaries

(I) Group	(J) Group	Dependent Variables' Significance Levels (p)		
		Effectiveness	Efforts	False Positives
Perspective	Ad hoc	0.039*	0.674	0.098
	Checklist	0.016*	0.078	0.199
Ad hoc	Perspective	0.039*	0.674	0.098
	Checklist	0.914	0.013*	0.917
Checklist	Perspective	0.016*	0.078	0.199
	Ad hoc	0.914	0.013*	0.917

* The mean difference is significant at the .05 level.

From table 4, it can be vividly inferred that there are truly significant differences between the effectiveness of Perspective and Ad hoc (p = 0.036 < 0.05), Perspective and Checklist (p = 0.016 < 0.05) inspection reading techniques. Between Ad hoc and others, there is a significant difference between Ad hoc and perspective effectiveness alone (p = 0.039 < 0.05). And in the case of Checklist, a significant difference is obtained between it and Perspective alone (p = 0.016 < 0.05).

Duncan's Post Hoc test in Table 5 shows that Mean Effectiveness value (31.25) is greater than the effectiveness of the other variables – Checklist and Ad hoc. This made us to conclude that Perspective – based reading techniques performed higher than both Checklist and Ad hoc reading techniques.

Efforts in terms of time in minutes spent by the reviewers in the inspection exercise, is only significant for Ad hoc and Checklist as well as between Checklist and Ad hoc.

Table 5: Duncan Post Hoc Test for Effectiveness

	N	Subset for alpha = .05	
Group Var	1	2	1
Checklist	8	27.5000	
Adhoc	8	28.0000	
Perspective	8		31.2500
Sig.		0.689	1.000

Means for groups in homogeneous subsets are displayed.
a Uses Harmonic Mean Sample Size = 8.000.

Table 6: Duncan post hoc test for effort

	N	Subset for alpha = .05	
Group Var	1	2	1
Adhoc	8	49.7500	
Perspective	8	53.7500	
Checklist	8		64.5000
Sig.		.403	1.000

Means for groups in homogeneous subsets are displayed.
a Uses Harmonic Mean Sample Size = 8.000.

Duncan post hoc test on Efforts shows that Checklist-based reading technique actually has the highest mean (64.5), which makes checklist group to actually expended more time than the Ad hoc group in the inspection exercise.

Results from table 4 shows that false positives are not truly significantly different for all the variables – Perspective, Ad hoc and Checklist based inspection artifact reading techniques.

5. Discussion of Results

Software inspection is a successful method for detecting faults in documents and codes produced in software development. Checklist-based and Ad hoc are the earlier reading techniques usually employed to detect errors in software artifacts. Perspective-based reading, proposed by Basili *et al.*, [3] in which a software product is inspected from the perspective of different stakeholders (analysts, designers, programmers, and so on) was later introduced.

Defects detection effectiveness of these three reading techniques was studied in this work. The fact that perspective –based reading technique outperforms other techniques is obvious from the fact that the reviewers were given specific tasks to perform on the codes. For instance, an analyst will have to inspect the code to ascertain that it conforms to the requirements specification and nothing more. Checklist-based reviewers were assisted with some checklists which gave some precise questions on what to look for in the code artifact. Therefore they are expected to perform higher than the Ad hoc group. However, this is not the case in this work.

Our results are in consonance with some related works in the literatures. To mention a few, Basili et al., [3] results show that Perspective-based reading technique is more effective than Ad-hoc or Checklist-based reading techniques. Giedre *et al.,* [6] results from their experiment to compare checklist based reading and perspective-based reading for UML design documents inspection shows that Checklist-based reading (CBR) uncovers 70% in defect detection while Perspective–based reading (PBR) uncovers 69% and that checklist takes more time (effort) than PBR. They also showed that that checklist-based consumes more time than Perspective-based technique. Porter and Votta [11] on their experiment for comparing defect detection methods for software requirements inspections show that checklist reviewers were no more effective than Ad hoc reviewers. Filippo and Giuseppe [5] on their work on evaluating defect detection techniques for software requirements inspections, also show that no difference was found between inspection teams applying Ad hoc or Checklist reading with respect to the percentage of discovered defects. Les [9] in his work on “Testing the value of checklists in code inspections” shows there is no evidence that checklists significantly improve inspections. Akinola and Osofisan [2] show that there is no statistical relationship between the false positives of Ad hoc and Checklist-based defect detection techniques.

6. Conclusion

Software inspection is very important in the software quality assurance. In this study, the statistical significant relationships among the effectiveness of Perspective-based, Checklist-based and Adhoc defect detection methods on software artifact inspections was questioned. It is concluded from this study that perspective based reading technique is a best choice for code inspection exercise. However, we look forward to authenticate our results with automated tools in the nearest future.

References

1. Abdusalam F. Ahmed Nwesri and Rodina Ahmad, (2000). An Asynchronous Software Inspection Model, *Malaysian Journal of Computer Science*, Vol. 13 No. 1, June 2000, pp. 17-26.
2. Akinola S.O. and Osofisan A.O, (2009). An empirical Comparative Study of Adhoc and Checklist Code Reading Techniques in Distributed Groupware Environment, *International Journal of Computer Science and Information Security (IJSCSIS)*, Vol. 5, No. 1, 2009.
3. Basili, V., Green, S., Laitenberger, O., Lanubile, F., Shull, F., Sorumgard, S., and Zelkowitz, M., (1996). The Empirical Investigation of Perspective-based Reading. Journal of Empirical Software Engineering, 2(1):133-164.
4. Capers Jones (2008). Applied Software Measurement, 3rd Ed. McGraw Hill.

5. Filippo Lanubile and Giuseppe Visaggio (2000), Evaluating defect Detection Techniques for Software Requirements Inspections, http://citeseer.ist.psu.edu/Lanubile00evaluating.html, downloaded Feb. 2010.
6. Giedre Sabaliauskaite, Fumikazu Matsukawa, Shinji Kusumoto, Katsuro Inoue, (2002). "An Experimental Comparison of Checklist-Based Reading and Perspective-Based Reading for UML Design Document Inspection," *ISESE*, p. 148, 2002 *International Symposium on Empirical Software Engineering* (ISESE'02).
7. IEEE Standard 1028-1997 (1998). *Standard for Software Reviews*. The Institute of Electrical and Electronics Engineering, Inc. ISBN 1-55937-987-1.
8. Laitenberger Oliver (2002), A Survey of Software Inspection Technologies, *Handbook on Software Engineering and Knowledge Engineering*, vol. II, 2002.
9. Les Hatton (2008). Testing the Value of Checklists in Code Inspections, *IEEE Software*, 25:4, July 2008, pp. 82 -88
10. Michael E. Fagan, (1976). Design and code inspections to reduce errors in program development. *IBM Systems Journal*, 15(3):182-211.
11. Porter, A. A. and Votta, L. (1998). Comparing Detection Methods for Software Requirements Inspection: A Replication using Professional Subjects. *Journal of Empirical Software Engineering*, vol. 3, no. 4, page 355-378.
12. Rifkin, S. and Deimel, L., (1994). Applying Program Comprehension Techniques to Improve Software Inspection. *Proceedings of the 19th Annual NASA Software Eng. Laboratory Workshop.NASA.*
13. Wheeler, D. A. and Brykczynski, B., (1996). Software Inspection: An Industry Best Practice, IEEE CS Press.
14. Porter, A. A., Votta, L. G. and Basili, V. R. (1995). Comparing detection methods for software requirements inspections: A replicated experiment. *IEEE Trans. on Software Engineering*, 21(Harvey, 1996):563-575.
15. Porter, A. A., Siy, H., P., Toman, C. A. and Votta, L. G. (1997). An Experiment to Assess the Cost-Benefits of Code Inspections in Large Scale Software development, *IEEE Transactions on Software Engineering*, vol. 23, No. 6, pp. 329 – 346.

Olalekan Akinola is a lecturer of Computer Science at the University of Ibadan, Nigeria. He had his PhD Degree in Software Engineering from the same University in Nigeria. He is currently working on Software Process Improvement modelling for software industry.

Ipeayeda Funmilola Wumi finished her Masters degree in Computer Science at the University of Ibadan in 2010. This work was actually part of her Masters Thesis.

IJCSIS REVIEWERS’ LIST

Mr. Masoud Rafighi, Islamic Azad University, Iran

Assist. Prof. RamBabu Pemula, NIMRA College of Engineering & Technology, India

Assist. Prof. Ms Rita Chhikara, ITM University, Gurgaon, India

Mr. Sandeep Maan, Government Post Graduate College, India

Prof. Dr. S. Muralidharan, Mepco Schlenk Engineering College, India

Associate Prof. T.V.Sai Krishna, QIS College of Engineering and Technology, India

Mr. R. Balu, Bharathiar University, Coimbatore, India

Assist. Prof. Shekhar. R, Dr.SM College of Engineering, India

Prof. P. Senthilkumar, Vivekanandha Institue of Engineering And Techology For Woman, India

Mr. M. Kamarajan, PSNA College of Engineering & Technology, India

Dr. Angajala Srinivasa Rao, Jawaharlal Nehru Technical University, India

Assist. Prof. C. Venkatesh, A.I.T.S, Rajampet, India

Mr. Afshin Rezakhani Roozbahani, Ayatollah Boroujerdi University, Iran

Mr. Laxmi chand, SCTL, Noida, India

Dr. Dr. Abdul Hannan, Vivekanand College, Aurangabad

Prof. Mahesh Panchal, KITRC, Gujarat

Dr. A. Subramani, K.S.R. College of Engineering, Tiruchengode

Assist. Prof. Prakash M, Rajalakshmi Engineering College, Chennai, India

Assist. Prof. Akhilesh K Sharma, Sir Padampat Singhania University, India

Ms. Varsha Sahni, Guru Nanak Dev Engineering College, Ludhiana, India

Associate Prof. Trilochan Rout, NM Institute Of Engineering And Technlogy, India

Mr. Srikanta Kumar Mohapatra, NMIET, Orissa, India

Mr. Waqas Haider Bangyal, Iqra University Islamabad, Pakistan

Dr. S. Vijayaragavan, Christ College of Engineering and Technology, Pondicherry, India

Prof. Elboukhari Mohamed, University Mohammed First, Oujda, Morocco

Dr. Muhammad Asif Khan, King Faisal University, Saudi Arabia

Dr. Nagy Ramadan Darwish Omran, Cairo University, Egypt.

Assistant Prof. Anand Nayyar, KCL Institute of Management and Technology, India

Mr. G. Premsankar, Ericcson, India

Assist. Prof. T. Hemalatha, VELS University, India

Prof. Tejaswini Apte, University of Pune, India

Dr. Edmund Ng Giap Weng, Universiti Malaysia Sarawak, Malaysia

Mr. Mahdi Nouri, Iran University of Science and Technology, Iran

Associate Prof. S. Asif Hussain, Annamacharya Institute of technology & Sciences, India

CALL FOR PAPERS
International Journal of Computer Science and Information Security

January - December
IJCSIS 2012
ISSN: 1947-5500

http://sites.google.com/site/ijcsis/

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc,), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on
its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embeded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC’s, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at http://sites.google.com/site/ijcsis/authors-notes .

ISSN 1947 5500